Language Universals and Misidentification: A Two-way Street

Iris Berent and Tracy Lennertz
Northeastern University, USA

Evan Balaban
McGill University, Canada and SISSA, Italy

Abstract
Certain ill-formed phonological structures are systematically under-represented across languages and misidentified by human listeners. It is currently unclear whether this results from grammatical phonological knowledge that actively recodes ill-formed structures, or from difficulty with their phonetic encoding. To examine this question, we gauge the effect of two types of tasks on the identification of onset clusters that are unattested in an individual's language. One type calls attention to global phonological structure by eliciting a syllable count (e.g., does *medif* include one syllable or two?). A second set of tasks promotes attention to local phonetic detail by requiring the detection of specific segments (e.g., does *medif* include an *e*?). Results from five experiments show that, when participants attend to global phonological structure, ill-formed onsets are misidentified (e.g., *mdif→medif*) relative to better-formed ones (e.g., *mlif*). In contrast, when people attend to local phonetic detail, they identify ill-formed onsets as well as better-formed ones, and they are highly sensitive to non-distinctive phonetic cues. These findings suggest that misidentifications reflect active recoding based on broad phonological knowledge, rather than passive failures to extract acoustic surface forms. Although the perceptual interface could shape such knowledge, the relationship between language and misidentification is a two-way street.

Introduction

The link between the distribution of phonological structures across languages and their representation by individual speakers is a significant discovery of modern linguistics (Jakobson, 1968; Prince

Corresponding author:
Iris Berent, Department of Psychology, Northeastern University, 125 Nightingale Hall, 360 Huntington Ave., Boston, MA 02115-5000, USA
Email: i.berent@neu.edu

& Smolensky, 1993/2004). Typological research (Greenberg, 1978) shows that certain phonological structures (e.g., the syllable *lba*) are systematically under-represented relative to others (e.g., *bla*). Moreover, structures that are under-represented across languages are harder for individual speakers to identify (Blevins, 2004). Although the misidentification of ill-formed structures is most striking when the relevant structures are unattested in a speaker's language, misidentification is not simply due to unfamiliarity. Indeed, ill-formed structures are harder to process than better-formed ones even when both are unattested in one's language. Such structures are systematically misidentified in various experimental tasks (e.g., *lba* is misidentified as *leba*, Berent, Steriade, Lennertz, & Vaknin, 2007; see also Davidson, 2006; Moreton, 2002; Zuraw, 2007) and they are harder to learn (Becker, Ketrez, & Nevins, 2011; Moreton, 2008; Wilson, 2006). The convergence between the typological tendencies (Hyman, 2008) and the behavior of individual speakers suggests that certain sound-structures are universally dispreferred to others.

Although the existence of phonological universals is well established, their source is contentious. One view attributes language universals to grammatical phonological knowledge that actively disfavors certain structures, rendering them ill-formed (de Lacy, 2006; Prince & Smolensky, 1993/2004). An alternative explanation proposes that the difficulties of processing ill-formed structures are the cause, rather than the consequence, of their dispreference. In this view, ill-formed structures may be difficult to accurately encode from acoustic cues (e.g., due to forward and backward masking, Moore, 2003), leaving them vulnerable to "innocent misperceptions" (Blevins, 2006; Ohala, 1990). Consequently, ill-formed structures undergo change in language evolution, and they would be difficult to acquire and process in experimental tasks. In this view, language universals are not actively represented in the human brain, but rather emerge from generic properties of human perception and production systems (Blevins, 2004, 2006).

Note that these two possibilities are not mutually exclusive—grammatical phonological universals could be actively represented in specialized human brain circuits, but could also be shaped, in part, by properties of more basic perceptual and articulatory interfaces (Hayes & Steriade, 2004). The question is not whether ill-formed structures can trigger "innocent misperceptions", but rather whether "innocent misperceptions" capture misidentifications fully.

Several authors have equated misidentifications with "innocent misperception", a tradition dating back to Baudouin de Courtenay (1845–1929), who underscores "the importance of errors in hearing (lapsus auris) when one word is mistaken for another, as a factor of change at any given moment of linguistic intercourse …" (cited in Blevins, 2007, p. 144). More recently, Peperkamp and Dupoux concluded that "Phonetic decoding, then, acts as a filter, in that many fine-grained acoustic details of speech sounds are lost as these sounds are mapped onto phonetic categories. […] Crucially, we argue that phonetic decoding equally accounts for suprasegmental 'deafnesses' and 'deafnesses' due to phonotactic constraints (cf. 2b-c), as well as to the corresponding repairs in loanword adaptations (cf. 1b-d)" (2003, pp. 368–369; see also Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011).

On an alternative hypothesis, misidentification reflects the active recoding of inputs, based on grammatical phonological knowledge. To articulate the contrast between these views, consider the processing of an ill-formed structure that is unattested in a speaker's language (e.g., *lbif*, represented by English speakers). Both accounts assume that people represent auditory linguistic stimuli at multiple levels, and that the final representation of ill-formed inputs by non-native speakers is typically distorted (relative to their representation by native speakers of a language where the structure is attested, e.g., Russian). Poor encoding predicts that given the Russian input *lbif*, English speakers would fail to register the occurrence of a consonant cluster in the surface phonetic form. Active recoding assumes grammatical phonological knowledge that actively recodes ill-formed inputs (e.g., *lbif*) as better-formed ones. In this view, English speakers possess at least two

Active phonological repair   **Passive failure of phonetic encoding**
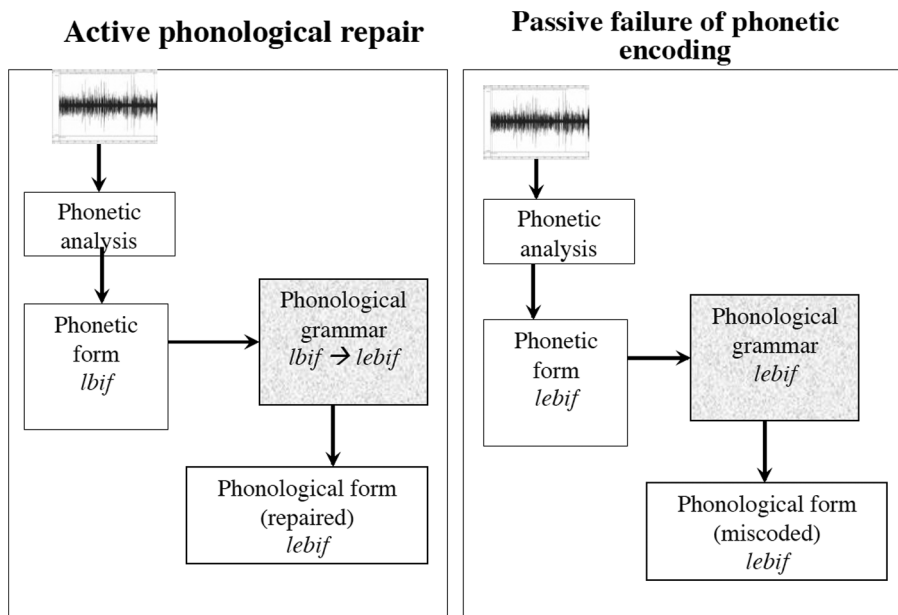
Figure 1. A schematic depiction of two rivalry accounts of the misidentification of ill-formed clusters: active phonological repair vs. passive failure of phonetic encoding. The locus of distortion is highlighted (see text for details).

representations of the input: a phonetic form that faithfully encodes the acoustic input as *lbif*—a representation isomorphic to the phonetic form extracted by a native (e.g., Russian) speaker—and a phonological form that recodes it based on phonological constraints (e.g., *lebif*). Misidentifications occur because hearers typically attend to the final phonological output, rather than the surface phonetic form (see Figure 1).

These two accounts differ on their predictions with respect to the scope of misidentifications. The view of misidentification as a passive inability to extract the surface phonetic forms of ill-formed inputs predicts that misidentifications should persist regardless of whether a task calls attention to phonetic properties of the input or not. In contrast, the account of active repair predicts that people possess a precise phonetic representation of inputs that they typically misclassify. Accordingly, conditions that encourage inspection of the surface phonetic forms might promote more accurate responses to ill-formed inputs (e.g., of *lbif*), responses that are as accurate as those given to better-formed ones (e.g., *bla*). The present experiments examine these predictions.

Our investigation builds on previous research examining the effect of ill-formedness on the perception of syllables that are all unattested in one's language (Berent et al., 2007). Ill-formedness here is defined by the structure of the onset cluster (the sequence of consonants at the beginning of the syllable), and it specifically concerns their sonority profile. Sonority (s) is a scalar phonological property that correlates with the intensity of segments (Parker, 2008). Obstruents (e.g., *d*, with a sonority level of 1, $s = 1$) are the least sonorous, followed by nasals (e.g., *m*, $s = 2$), liquids (e.g., *l*, $s = 3$) and glides (e.g., *w*, $s = 4$). Accordingly, onsets such as *bl* and *ml* manifest a rise in sonority ($\Delta s = 2$ and $\Delta s = 1$, respectively), whereas *lb* and *md* manifest sonority falls ($\Delta s = -2$ and $\Delta s = -1$, respectively). Linguists have long noticed that onsets with large sonority distances (e.g., sonority rises) are preferred to onsets with smaller distances (e.g., sonority falls, Clements, 1990; Smolensky, 2006). Such

preferences are seen in the distribution of these onsets across languages and their perception by individual speakers. Concerning the typology, an inspection of 90 languages (Greenberg, 1978, analyzed in Berent et al., 2007) suggests that small sonority distances are less frequent than larger ones, and that the presence of small sonority distances in any given language implies the presence of larger distances. Onsets with small sonority distances (positive or negative) also appear to be systematically misidentified by speakers of various languages. For example, people are more likely to misidentify onsets with sonority falls (e.g., *lbif* is perceived as *lebif*) relative to better-formed onsets of rising sonority (e.g., *bnif*) even when both types are absent in their language, as is the case for English (Berent et al., 2007) and even when their language lacks syllables beginning with *any* consonant-sequence, as is the case of Korean (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008).

Of specific interest here is why onsets of falling sonority are misidentified—do people fail to extract their phonetic forms, or do they actively recode them? To examine this question, the following experiments compare two types of nasal-initial onsets: one with sonority rises (*ml*) and one of falling sonority (*md*). We examine the identification of these onsets under two conditions—either ones that call attention to global phonological structure by eliciting a syllable count (e.g., does /mədɪf/ include one syllable or two?), or conditions that promote attention to local phonetic detail by requiring the detection of specific segments (e.g., the presence of a schwa in /mədɪf/). In view of past results using the syllable-count procedures, we expect onsets of falling sonority to be misidentified as disyllabic. If misidentifications reflect a passive failure to encode the surface phonetic forms of ill-formed onsets, then such difficulties should persist under conditions that call attention to phonetic form. Although the heightened attention could conceivably improve overall accuracy, the disadvantage of onsets of falling sonority relative to the better-formed onsets of rising sonority should persist. In contrast, if repair is due to active phonological recoding, and if, further, the phonetic form of such onsets is both accurately encoded and accessible then, once participants attend to the phonetic form, the difficulty in the identification of ill-formed onsets should be greatly reduced, and possibly, even eliminated. Moreover, the shift in processing mode should be further evident as an increased sensitivity to non-distinctive phonetic detail. In what follows, we examine these predictions in two sets of experiments. In Part 1, participants are presented with continua in which the duration of the pretonic vowel is gradually manipulated, ranging from a disyllabic form (e.g., /mədɪf/) to its monosyllabic counterpart (e.g., /mdɪf/). In Part 2, participants are presented only with the two endpoints—monosyllables or disyllables.

2 Part 1: Processing CCVC-CəCVC continua

Experiments 1 and 2 compare the identification of two types of continua. Each such continuum was generated by a procedure of incremental splicing along the lines described by Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999). We first had a native English speaker naturally produce the disyllabic counterparts (e.g., *melif* and *medif*), and next continuously excised the pretonic vowel *e* in five increments. This procedure yielded a continuum of six steps, ranging from the original disyllabic form to a monosyllabic form with an onset cluster (e.g., *mlif*, *mdif*).

Generally speaking, we expect that, as the phonetic duration of the pretonic vowel increases, people should be more likely to categorize the item as disyllabic. Moreover, responses to items at the monosyllabic endpoint of the continuum should be typically modulated by their sonority profile: *md*-type onsets should be more likely to be identified as disyllabic relative to their *ml*-type counterparts, replicating previous results with nasal clusters (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Balaban, Lennertz, & Vaknin-Nusbaum, 2010). In what follows,

we consider such disyllabic responses to fully-monosyllabic inputs as misidentification. To gauge the source of misidentification—whether it is due to passive encoding failure or active recoding—we compare the identification of such onsets using two tasks. In each task, participants hear one item at a time and classify it on a dimension that correlates with its syllable structure. The two tasks differ on the dimension of classification—one that calls attention either to global phonological structure (the number of syllables) or to fine-grained phonetic form (the presence of a pretonic vowel between the initial two consonants). If the misidentification of ill-formed onsets of falling sonority reflects a passive failure, then the difference in task demands might change the overall accuracy of *mdif* items, but not their disadvantage relative to *mlif* items. In contrast, if misidentification reflects an active modification of a precise phonetic form, and if this form remains accessible, then tasks that encourage attention to phonetic properties should eliminate the disadvantage of onsets of falling sonority, yielding comparable results for *mdif*- and *mlif*-type items, coupled with greater sensitivity to non-distinctive phonetic detail.

2.1 Experiment 1: Beat count

Experiment 1 seeks to replicate the typical misidentification of ill-formed onsets using a task that underscores their global phonological properties (their syllable count). To reduce meta-linguistic strategies, we asked participants to count the number of "beats" in the stimulus (rather than syllables—a meta-linguistic label), and illustrated the task using existing words (e.g., *sport* contains one beat; *support* contains two). If the ill-formed *mdif*-type inputs are typically misidentified (as *medif*) relative to their better-formed *mlif*-type counterparts, then as the duration of the pretonic vowel decreases, participants should be more likely to classify *mdif*-type items as disyllabic (i.e., as containing two beats).

2.1.1 Method

2.1.1.1 Participants.
Twelve native English speakers, students at Florida Atlantic University, took part in this experiment in partial fulfillment of a course requirement.

2.1.1.2 Materials.
The materials consisted of a continuum of auditory forms, ranging from monosyllables to disyllables. The monosyllabic forms included two types of nasal-initial onsets, manifesting either a sonority rise or fall (e.g., /mlɪf/, /mdɪf/). These items were arranged in pairs, matched for their rhyme. There were three such pairs (/mlɪf/-/mdɪf/, /mlɛf/-/mdɛf/, /mlɛb/-/mdɛb/). Each pair member was generated by a procedure of incremental splicing, as described in Dupoux et al. (1999). We first had a native English speaker (naive to this research project) naturally produce the disyllabic counterparts (e.g., /məlɪf/-/mədɪf/) in a sentential context ("This is X"), and selected pairs that were matched for total length, intensity, the duration of the pretonic schwa (68 ms for both /məlɪf/-, /mədɪf/-type items; for an illustration, see Figure 2), and its fundamental frequency (for /məlɪf/-type items: $M = 198.68$ Hz, $SD = 2.19$; for /mədɪf/-type items: $M = 199.34$ Hz, $SD = 4.79$).[1] We next continuously extracted the pretonic vowel at the zero crossings in five increments, moving from its center outwards. This, in turn, yielded a continuum of six steps, ranging from the original disyllabic form to an onset cluster, in which the pretonic vowel was fully removed. The number of pitch periods in Stimuli 1–5 was 0, 2, 4, 6, and 8, respectively; Stimulus 6 (the original disyllable) ranged from 12 to 15 pitch periods.

Each of the three pairs was presented in all 6 durations, resulting in a block of 36 trials. Each such block was repeated four times, yielding a total of 144 trials. The order of trials within each block was randomized.
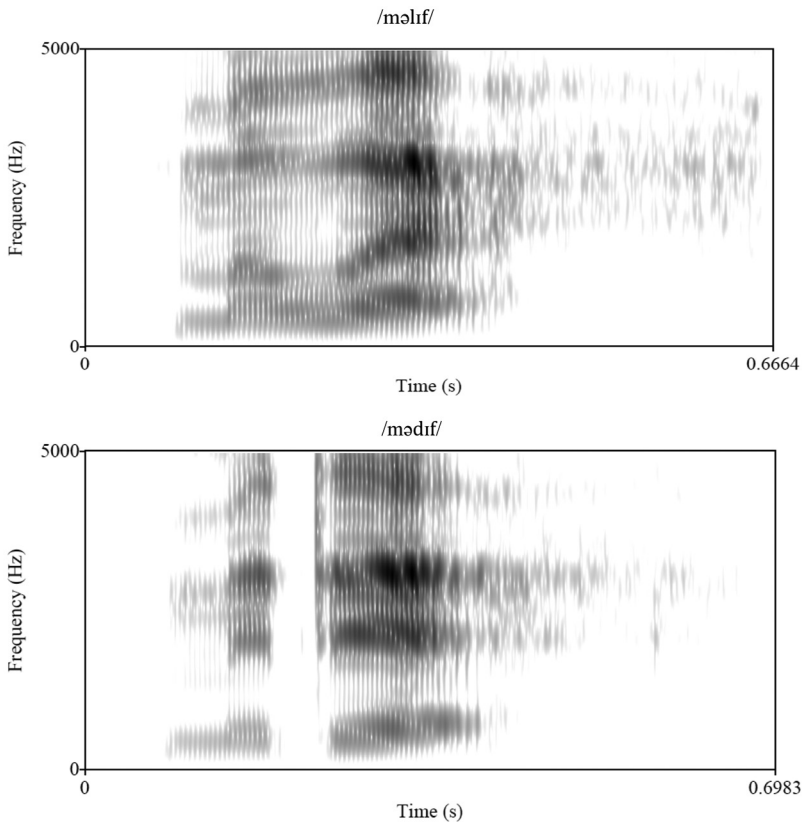
Figure 2. An illustration of the naturally-produced counterparts of onsets with rising and falling sonority (*melif* vs. *medif*).

2.1.1.3 Procedure.   Participants were seated in front of a computer wearing headphones. Each trial began with a message indicating the trial number. Participants initiated the trial by pressing the space bar, triggering the presentation of a fixation point (+, displayed for 500 ms), followed by the auditory stimulus. Participants were asked to indicate whether the stimulus contained one beat or two by pressing the appropriate key (1 = one beat; 2 = two beats). The task was illustrated using English words (e.g., *sport*, *support*, spoken naturally by the experimenter), and a brief practice with novel words, produced and spliced as the experimental materials.

2.1.2 Results and discussion.   Figure 3 plots the proportion of disyllabic ("two beat") responses as a function of the duration of the pretonic vowel and the type of the onset. As the duration of the pretonic vowel increased, people were more likely to classify the stimulus as disyllabic. But interestingly, disyllabic responses were more frequent for ill-formed onsets of falling sonority compared to better-formed sonority rises, except when the item was fully disyllabic.

These conclusions are supported by the significance of the interaction in a 6 vowel-duration × 2 onset-type ANOVA, $F(5, 55) = 10.78, p < .00001$. The simple main effect of vowel duration was significant for both sonority rises, $F(5, 55) = 51.74, p < .0002$, and falls, $F(5, 55) = 6.72, p < .0001$, indicating an increase in disyllabic responses with vowel duration. But importantly, disyllabic
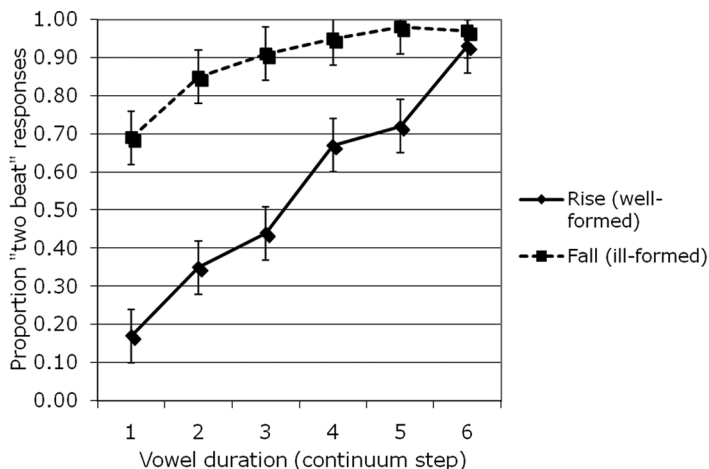
Figure 3. The proportion of "two beat" responses to well-formed onsets with sonority rises and ill-formed onsets with sonority falls as a function of vowel duration in Experiment 1. Vowel duration is defined along a six-step continuum (1 = monosyllabic; 6 = disyllabic). Error bars reflect 95% confidence intervals constructed for the difference between the means of sonority rises and falls.

responses were more likely for ill-formed items with sonority falls relative to rises, resulting in significant simple effects of onset type at each of the five initial steps (step 1: $F(1, 11) = 41.79$, $p < .0002$; step 2: $F(1, 11) = 40.89$, $p < .0002$; step 3: $F(1, 11) = 34.03$, $p < .0002$; step 4: $F(1, 11) = 21.99$, $p < .0008$; step 5: $F(1, 11) = 22.14$, $p < .0007$). In contrast, when participants were presented with fully disyllabic forms, in step 6, responses to the two types of items did not differ reliably, $F(1, 11) = 3.76$, $p < .08$.

2.2 Experiment 2: Beat count vs. vowel detection

The results of Experiment 1 confirm that ill-formed onsets of falling sonority tend to be misidentified. This effect of ill-formedness on identification is akin to an effect of vowel duration. For example, the rate of disyllabic misidentification of monosyllables with sonority falls (at step 1) is comparable to the rate of disyllabic responses to sonority rises with a substantial pretonic vowel, at step 4. The misidentification of onsets of falling sonority cannot be due to the inherent disyllabicity of our experimental materials, as Russian speakers categorize the very same stimuli as monosyllabic (Berent et al., 2010; Experiment 3). This contrast suggests either that English speakers are effectively oblivious to the phonetic cues that distinguish *mdif-* and *medif-*type forms, or that disyllabic responses occur because people typically consult a representation that actively recodes the input in accordance with their phonological knowledge.

Experiment 2 was designed to adjudicate between these possibilities by comparing the perception of the same materials using two tasks. One is the beat-count task, employed in Experiment 1. In a second vowel-detection task, participants monitored the presence of a vowel between the two initial consonants. Because beat counting encourages attention to global phonological structure, we expected it to yield a greater rate of disyllabic responses for onsets of falling sonority. Of interest are the results of vowel detection—a task that calls attention to local, fine-grained phonetic properties of the critical vowel. If misidentification entails a passive inability to encode the phonetic form, then both tasks should yield a higher rate of disyllabic responses for ill-formed onsets

relative to better-formed ones. Although attention to the phonetic form could conceivably improve overall performance, it should not eliminate the relative disadvantage of sonority falls in the vowel-detection task. In contrast, if the phonetic form of sonority falls is precisely encoded and accessible, then the vowel-detection task should reduce the disadvantage of ill-formed onsets of falling sonority.

The change in processing mode, moving attention from global phonological structure (in beat-count) to fine-grained phonetic detail (in vowel-detection), should be further evident as an increased sensitivity to non-contrastive phonetic cues. Recall that our vowel continua presented in steps 1–5 were generated by a procedure of incremental splicing, whereas the sixth step was produced naturally. Because splicing generates discontinuity, its presence in steps 1–5 should signal bifurcation—a cue for disyllabicity, whereas its absence in the unspliced endpoint might attenuate the disyllabic response. Our past research with these continua (Berent et al., 2010) has suggested that participants are highly sensitive to this cue, especially when they are unfamiliar with the phonetic properties of the input (e.g., when the speech signal is transformed to appear as non-speech, or when participants are non-native speakers of English). To the extent that the vowel-detection task indeed promotes attention to phonetic cues, then we expect participants to be likewise sensitive to the phonetic continuity of the unspliced endpoint. This, in turn, would result in a reduction in the proportion of disyllabic identification of the unspliced endpoints in the vowel-detection task compared to the beat-count procedure.

2.2.1 Method

2.2.1.1 Participants.   Twenty-four native English speakers, students at Florida Atlantic University, took part in this experiment in partial fulfillment of a course requirement.

2.2.1.2 Materials.   The materials were as in Experiment 1.

2.2.1.3 Procedure.   Each participant took part in two tasks: beat-count and vowel-detection. The beat-count was the same task used in Experiment 1; the vowel-detection task was identical, except that participants were now asked to determine whether the stimulus "contained a vowel between the first two sounds". These tasks were presented in separate experimental blocks, with order counter-balanced. Prior to the experiment, participants were informed that the experiment contains two parts, with different tasks and instructions (pilot work suggests this warning increased phonetic vigilance even before the task had shifted), but the instructions for each task were only given at the beginning of the relevant block. As in Experiment 1, each experimental block began with an illustration of the task using real English words and a brief practice with spliced novel words. The experimental block for each task included two repetitions of the 36 stimuli (3 stimulus pairs × 2 onset type × 6 vowel duration), a total of 72 trials per block.

2.2.2 Results and discussion.   Figure 4 plots the proportion of disyllabic responses for sonority rises and falls as a function of vowel length and the task—beat-count vs. vowel-detection. An inspection of the means suggests that, as the duration of the vowel increased, participants were more likely to consider the target as disyllabic. Likewise, given items at the monosyllabic end, disyllabic responses were more likely for sonority falls compared to rises. However, the misidentification of ill-formed onsets of falling sonority was far more pronounced in the beat-count compared to the vowel-detection task.

A 2 order × 2 task × 2 onset-type × 6 vowel-duration ANOVA yielded a reliable task × onset type × vowel interaction, $F(5, 110) = 3.94$, $MSE = .0272$, $p < .003$, which was not further
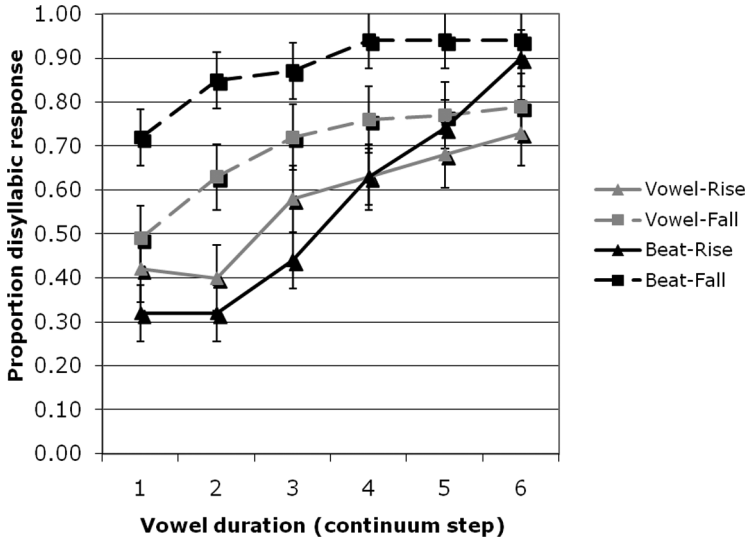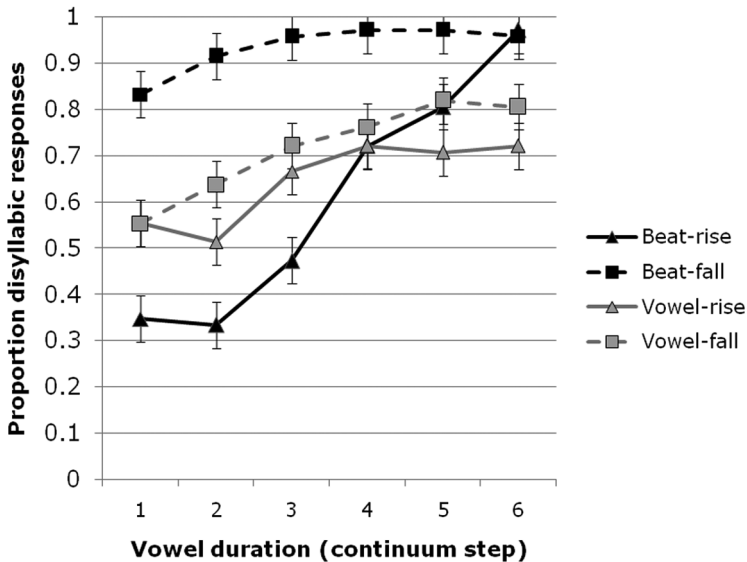
Figure 4. The proportion of disyllabic responses as a function of task, onset type and vowel duration in Experiment 2. Vowel duration is defined along a six-step continuum (1 = monosyllabic; 6 = disyllabic). Error bars reflect 95% confidence intervals constructed for the difference between the means of sonority rises and falls. Note: vowel = vowel detection; beat = beat counting; rise = sonority rise; fall = sonority fall.

modulated by the order of presentation (for the four-way interaction, $F < 1$). To investigate this interaction, we first examined the effect of onset type and vowel duration for each task separately using 2 onset-type × 6 vowel-duration ANOVAs. The critical vowel-duration × onset-type interaction was significant in the beat-count task, $F(5, 115) = 13.10$, $p < .0001$, but not in the vowel-detection procedure, $F(5, 115) = 1.43$, $p < .22$. Tukey HSD tests of the effect of onset type in the beat-detection task further confirmed that, compared to sonority rises, sonority falls were more likely to elicit disyllabic responses in steps 1–5 (all p's < .002), but not in the final step ($p > .99$).

To demonstrate that the experimental task specifically affected responses to unambiguously monosyllabic inputs of falling sonority, we next examined the effect of task for step 1, separately. A 2 order × 2 task × 2 onset-type ANOVA yielded a reliable interaction of task × onset type, $F(1, 22) = 11.07$, $MSE = .0596$, $p < .004$, which was not further modulated by the order of presentation ($F < 1$). Tukey HSD tests confirmed that the two tasks differed reliably for sonority falls ($p < .02$), but not for sonority rises ($p > .48$, n.s.). While in the beat-count task, sonority falls were reliably more likely to be misidentified as disyllabic compared to sonority rises ($p < .0003$), in the vowel-detection task, disyllabic responses did not differ reliably for onsets of rising and falling sonority ($p > .76$, n.s.).

Finally, to assure that the insensitivity of the vowel-detection task to onset type is not due to fatigue or carry-over effects from the performance of the beat-count task, in the previous block of trials, we next compared the results of the two tasks when they were each presented in the first block (see Figure 5). A 2 task × 2 onset-type × 6 vowel-duration ANOVA yielded a reliable three-way interaction, $F(5, 110) = 7.05$, $MSE = .025$, $p < .0001$. Additional 2 onset-type × 6 vowel-duration ANOVAs, performed separately on each task confirmed that the interaction was highly significant in the beat-count task, $F(5, 55) = 13.63$, $MSE = .0231$, $p < .0001$. In contrast, for the vowel-detection task, the ANOVA yielded only a significant main effect of vowel duration,

Figure 5. The proportion of disyllabic responses at the first block of trials in Experiment 2 as a function of task, onset type and vowel duration. Vowel duration is defined along a six-step continuum (1 = monosyllabic; 6 = disyllabic). Error bars reflect 95% confidence intervals constructed for the difference between the means of sonority rises and falls. Note: vowel = vowel detection; beat = beat counting; rise = sonority rise; fall = sonority fall.

$F(5, 55) = 2.80$, $MSE = .076$, $p < .03$. Neither the effect of onset type, $F(1, 11) = 1.15$, $p < .31$, nor its interaction with vowel duration ($F < 1$) approached significance.

The ability of participants in the vowel-detection task to accurately encode the phonetic forms of ill-formed onsets of falling sonority suggests that their misidentification in the beat-count procedure is not due to a principled inability to encode the phonetic form of such onsets. Further evidence that the differences between the outcomes of the two tasks is, in fact, due to differences in the attention to fine-grained phonetic detail is presented by responses to the disyllabic endpoints. Recall that, unlike steps 1–5, the sixth endpoint was unspliced, hence, it manifested greater phonetic continuity—a cue that conflicts with the expected disyllabic responses. If participants in the vowel-detection task are indeed tuned to phonetic cues, then they should also be less likely to interpret the unspliced disyllabic endpoint as disyllabic compared to the beat-detection procedure. A 2 order × 2 task × 2 onset-type ANOVA on the unspliced disyllabic endpoint indeed yielded a reliable main effect of task, $F(1, 22) = 5.39$, $MSE = .109$, $p < .03$, which was not further modulated by onset type or order (all F's < 1). The attenuation in the disyllabic identification of the unspliced endpoints presents further evidence that vowel detection promoted a qualitative shift in performance that enhanced participants' sensitivity to phonetic detail. Crucially, once participants attended to the phonetic detail, the difficulty in the processing of sonority falls was eliminated. This finding is inconsistent with the possibility that the misidentification of sonority falls (e.g., in Experiment 1) is due to an inability to encode their phonetic form.

The results from Experiments 1 and 2 are nonetheless limited in several respects. While the use of a psychophysical manipulation of schwa duration allows one to systematically gauge a phonetic dimension of interest, these gains come at a cost of reducing the number of items. Accordingly, one might worry that the inability of English participants to identify onsets of falling sonority could be

due to some idiosyncratic phonetic properties of a handful of items, rather than an inherent (phonetic or phonological) property of sonority falls in general. Another concern is that our continuous vowel-length manipulation rendered the items too similar, thereby artificially elevating misidentification due to confusion regarding the experimental task.

But there are several reasons to doubt item- and task-artifacts as the explanation for the results. The ability of participants in vowel detection to identify the very same onsets of falling sonority once they attended to their phonetic form demonstrates that these stimuli were not inherently flawed. Indeed, the misidentification of onsets of falling sonority replicates previous findings obtained with another set of nasal-initial items (Berent et al., 2009). Similarly, the systematic effect of vowel duration demonstrates that participants were quite sensitive to this dimension and followed the task quite well. Moreover, subsequent research showed that Russian participants identify monosyllables of falling sonority accurately using the very same materials and task (Berent et al., 2010). The systematic performance of English participants, on the one hand, and their divergence with Russian participants, on the other, questions these artifactual explanations for the findings. To further demonstrate the generality of our conclusions, Part 2 extends our investigation to a new set of materials and tasks.

3 Part 2: The role of phonetic cues in processing CCVC-CəCVC contrasts

The experiments presented in Part 2 further examine the identification of onsets of rising and falling sonority as compared with their disyllabic counterparts (e.g., *melif*, *medif*). Unlike the continuous vowel manipulation in Part 1, here we set a dichotic contrast between naturally-produced disyllables and their monosyllabic counterparts—items in which the pretonic schwa was entirely removed. Of interest is whether the identification of such onsets is modulated by participants' attention to phonetic detail.

As a baseline, Experiment 3 first compares the onsets of rising and falling sonority using a variant of the beat-count task. Participants in this experiment are asked to determine whether the input includes one syllable or two—a simplified version of the beat-count task, which likewise calls attention to global aspects of phonological structure. The dichotic contrast between monosyllables and disyllables, the simplification in the task and the use of a larger number of newly-recorded items are all designed to demonstrate that the previous results are not limited to a particular set of items or task demands.

Experiments 4 and 5 next seek to illuminate the basis for the typical misidentification of *md*-type onsets. To this end, we employ two distinct tasks that elicit attention to local phonetic detail. In each task, participants perform a forced choice as to whether a specific segment is present or absent. In Experiment 4, participants spot the presence of a pretonic schwa (a replication of the vowel-detection task in Experiment 2), whereas in Experiment 5 they monitor the second onset consonant (e.g., does the target *mdif* include a *d*?). If misidentification reflects a passive failure to encode the phonetic form, then the difficulties in identifying ill-formed onsets should persist irrespective of task demands. In contrast, the view of misidentification as an active recoding leads to three quite different predictions. Specifically, if participants possess a faithful phonetic record of ill-formed onsets, then once attention to their phonetic form is encouraged, (a) ill-formed onsets (e.g., *mdif*) should be processed more accurately, perhaps as accurately as their better-formed counterparts (e.g., *mlif*); (b) the identification of ill-formed monosyllables should differ from their disyllabic counterparts (e.g., *medif*); and (c) the improved identification of ill-formed onsets should be accompanied by enhanced sensitivity to non-distinctive phonetic cues, such as those related to coarticulation.

3.1 Experiment 3: Syllable count

3.1.1 Method

3.1.1.1 Participants.   Eighteen native English speakers, students at Northeastern University, took part in this experiment in partial fulfillment of a course requirement.

3.1.1.2 Materials.   The materials consisted of six pairs of CCVC monosyllables along with their disyllabic, CəCVC counterparts (see Appendix). Monosyllables all included onsets that are unattested in English, either sonority rises (*ml*) or sonority falls (*md*). Members of a monosyllabic pair were matched for their rhyme, and differed only in the structure of the onset (e.g., mlɪf, mdɪf). Disyllables were identical to the monosyllables, except for the presence of a schwa between the initial onset consonants (e.g., məlɪf, mədɪf).

The disyllables were recorded by a native English speaker (different from the one used in Experiment 1) in a randomized list. The disyllabic pair members, counterparts of the monosyllables of rising and falling sonority, were next selected to match for the duration of the pretonic schwa and their loudness was equated as closely as possible. The duration of the schwa and the second consonant (monitored in Experiments 4 and 5, respectively) is presented in Table 1.

We next generated the monosyllables by excising the pretonic schwa from their disyllabic counterparts at the point of the zero crossing (to eliminate acoustic artifacts). The resulting 6 pairs of monosyllables and their disyllabic counterparts (a total of 24 items) were presented to participants in four repeated blocks (a total of 192 trials), with the order of presentation randomized within a block. To illustrate the task, participants were first provided with a short practice session, consisting of four minimal pairs of English monosyllabic words and their disyllabic counterparts (e.g., *blow-below*).

3.1.1.3 Procedure.   Each trial was preceded by a message indicating the trial number. Participants initiated the trial by pressing the space bar, triggering the presentation of a fixation point (a plus sign, displayed for 250 ms) followed by an auditory stimulus. Participants were asked to determine whether they have heard one syllable or two, and indicate their response by pressing one of two keys (1 = one syllable; 2 = two syllables). Slow responses (responses longer than 2000 ms) triggered a computerized warning message ("too slow"). Participants did not receive feedback on their accuracy.

3.1.2 Results and discussion.   Mean response accuracy and response time as a function of onset type and the number of syllables is presented in Figure 6. As expected, the simplification of the task

Table 1. The duration of the experimental materials used in Experiments 3–6.

|	Rise		Fall	
	Mean (ms)	SD (ms)	Mean (ms)	SD (ms)
C1	118.44	12.30	113.03	6.30
Schwa	148.43	18.20	150.37	8.90
C2	128.17	10.80	119.43	20.60
Rhyme	486.53	12.10	535.49	55.20
Total duration (monosyllables)	733.14	28.60	767.95	46.70
Total duration (disyllables)	881.57	33.81	918.31	53.20
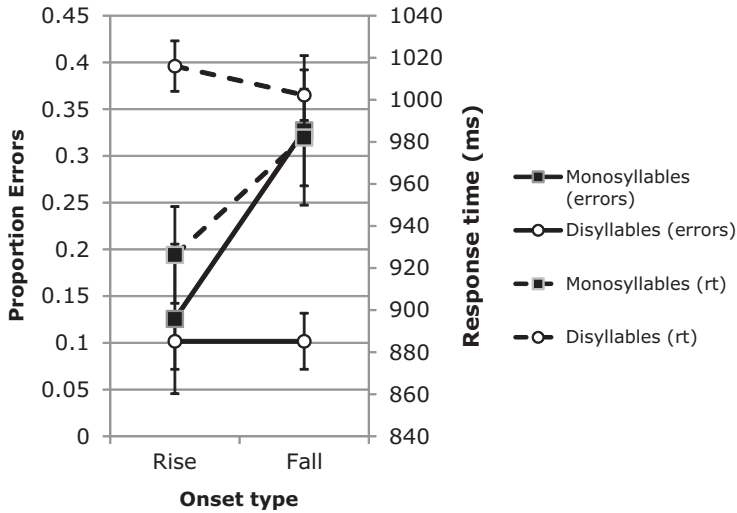
Figure 6. Response accuracy (% errors) and response time as a function of the structure of the monosyllabic counterpart and the number of syllables in Experiment 3 (syllable count). Error bars reflect 95% confidence intervals constructed for the difference between the means of sonority rises and falls.

elevated the level of accuracy. Accordingly, we analyzed performance in Experiments 3–6 using both accuracy and response time as dependent measures. In this and all subsequent experiments, we removed outliers (correct response falling 2.5 SD above the grand mean; less than 3% of the data in each experiment) from the analysis of response time. For viewing convenience, we plot response time and accuracy in a single figure. Note, however, that accuracy now corresponds to the proportion of *errors* (rather than correct responses).

An inspection of the means suggests that onset type modulated responses to monosyllables. A 2 onset-type ANOVA yielded significant effects in both response accuracy, $F(1, 17) = 6.72$, $MSE = .054$, $p < .02$, and response time, $F(1, 17) = 5.98$, $MSE = 4694$, $p < .03$. Similar analyses of responses to disyllabic items yielded no reliable effects (for response accuracy: $F < 1$; for response time: $F(1, 17) = 1.31$, $MSE = 1310$, $p > .26$, n.s.). Thus, replicating the findings of Experiment 1, as well as previous findings by Berent et al. (2009), monosyllables with ill-formed onsets of falling sonority were harder to identify as monosyllabic—they elicited slower and less accurate responses compared to monosyllables of rising sonority.

3.2 Experiment 4: Vowel detection

The findings from Experiment 3 establish that monosyllables of falling sonority are harder to identify, and that this phenomenon replicates across different materials and tasks. Experiments 4 and 5 now turn to investigate the source of this effect. If the difficulty with sonority falls only stems from an inability to encode their phonetic form, then their relative disadvantage should persist irrespective of task demands. Conversely, if misidentification reflects active recoding, and if, further, the phonetic form of such items is precisely encoded, then conditions that encourage attention to the phonetic form should allow participants to identify such onsets accurately—perhaps as well as their better-formed counterpart.
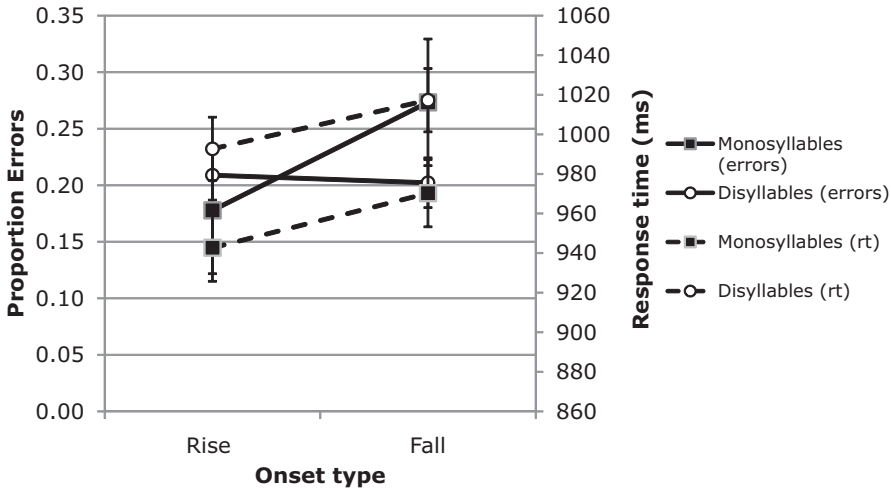
Figure 7. Response accuracy (% errors) and response time as a function of the structure of the monosyllabic counterpart and the number of syllables in Experiment 4 (schwa detection). Error bars reflect 95% confidence intervals constructed for the difference between the means of sonority rises and falls.

To examine this prediction, Experiment 4 introduces a slight modification of the task, analogous to the procedure used in Experiment 2. As in Experiment 3, participants effectively discriminated monosyllables from their disyllabic counterparts, but rather than doing so by inspecting the global phonological structure of the input—the number of syllables—participants were now asked to attend to a local phonetic cue—the presence of a pretonic schwa (e.g., does *medif* include an *e*?). If ill-formed inputs, such as *mdif*, are encoded accurately, then once participants attend to phonetic detail, the performance with such onsets should be comparable to that of their disyllabic counterparts.

3.2.1 Method

3.2.1.1 Participants.   Eighteen native English speakers, students at Northeastern University, took part in this experiment in partial fulfillment of course requirements.

The materials, design, the number of trials and procedure were the same as in Experiment 3, with the only change that participants were now asked to determine whether the auditory input included a vowel between the two initial consonants and indicate their responses by pressing the appropriate key (1 = *e* is present, 2 = *e* is absent).

3.2.2 Results and discussion.   Mean response accuracy (the proportion of errors) and response time are presented in Figure 7. A 2 onset-type ANOVA on the monosyllabic items did not find a reliable main effect of onset type in either response accuracy, $F(1, 17) = 2.89$, $MSE = .028$, $p < .11$, n.s., or response time, $F(1, 17) = 2.38$, $MSE = 2870$, $p < .15$. Thus, participants were no more likely to falsely state that a schwa was present in monosyllables of falling sonority (e.g., *mdif*) compared to sonority rises (e.g., *mlif*). Similarly, the structure of the monosyllabic counterpart did not reliably modulate the correct detection of the schwa in disyllabic stimuli (e.g., *medif* vs. *melif*; for response accuracy: $F < 1$; for response time: $F(1, 17) = 2.42$, $MSE = 22.53$, $p < .14$). Thus, once participants attended to the schwa—a local phonetic cue for disyllabicity—responses to monosyllables of falling sonority did not reliably differ from their rising-sonority counterparts.
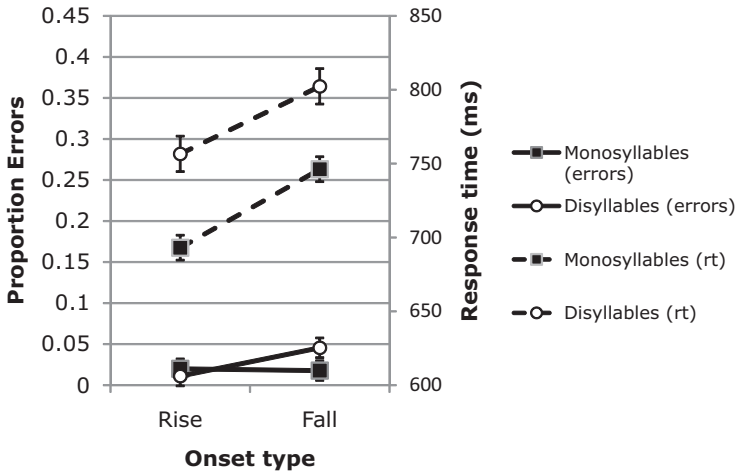
3.3 Experiment 5: C2 detection

The finding that participants are no more likely to falsely detect a schwa in *mdif* than in *mlif* is consistent with the possibility that the phonetic forms of these two structures are encoded with comparable accuracy. Moreover, because /mdɪf/ and /mədɪf/ contrast on their schwa, the effective encoding of this segment would suggest that the typical misidentification of *mdif* as disyllabic is not due to a passive encoding failure. Nonetheless, this finding does not rule out encoding failure entirely. Perhaps the greater disyllabic misidentification of *mdif* (relative to *mlif*) results from a failure to encode the second onset consonant. If participants mistake the *d* for pretonic schwa (e.g., /mdɪf/→/əmɪf/), then this failure, in turn, could give rise to a disyllabic response as well. Because vowel detection specifically calls for the detection of a schwa between the initial two consonants, participants could have correctly identified the absence of a schwa in sonority falls (in the vowel-detection tasks, Experiments 2 and 4) despite having distorted their phonetic form as disyllabic (e.g., as *emif*). To rule out this possibility, Experiment 5 specifically examines the ability to detect the second onset consonant.

Participants in Experiment 5 were asked to detect the second onset consonant across several counter-balanced blocks of trials. In one block, they determined whether or not the stimulus included the consonant *l*. In this condition, participants responded positively to *mlif* and *melif*, and negatively to *mdif* and *medif*. In another block of trials, participants monitored the presence of *d* (present in *mdif* and *medif*; absent in *mlif* and *melif*). If participants encode the phonetic form of *mdif* accurately, then their accuracy for detecting the presence of the target consonant should be similar for *mdif* and *medif*. Moreover, participants in this task should also exhibit sensitivity to the presence of non-distinctive phonetic cues. Several observations suggest that the phonetic cues for *l* are more robust than *d*. Liquids carry stronger internal and transitional cues (Wright, 2004), and the anticipatory resonances associated with the articulation of the liquid span numerous syllables (Heid & Hawkins, 2000; West, 1999). If participants in this experiment are tuned to such non-distinctive phonetic cues, then we expect target detection to be easier for *l* compared to *d*.

3.3.1 Method

3.3.1.1 Participants.   Eighteen native English participants took part in this experiment in partial fulfillment of a course requirement.

3.3.1.2 Materials and procedure.   The materials and procedure were the same as the ones used in Experiment 3, except that participants were now asked to monitor a consonant, either an *l* or a *d*. To this end, the materials were arranged in eight alternating blocks. In one block, participants were asked to monitor for an *l*, in another, they monitored for a *d*. Each block included 24 trials (a total of 192 trials per participant), and the order of blocks (monitor for *l* vs. *d*) was counter-balanced across participants. Each block was preceded by a colorful display, reminding participants of the task (e.g., "spot *d*"). To minimize confusion, the instructions for *l*- and *d*-blocks were each presented in different colors. Each trial was preceded by a message indicating the trial number and the target to-be-spotted (e.g., "spot *l*"). Participants initiated the trial by pressing the space bar, triggering the presentation of a fixation point (a plus sign, displayed for 250 ms) followed by an auditory stimulus. Participants were asked to indicate whether or not the auditory stimulus began with the designated target by pressing one of two keys (1 = target present; 2 = target absent). Slow responses (responses slower than 1500 ms) triggered a computerized warning message ("too slow"). At the end of each block, participants received a computerized message informing them of their accuracy in the previous block of trials. Each block was preceded by a short practice, consisting of four items that did not form part of the experimental materials (e.g., *mlig*, *mdesh*, *medig*, *melesh*).

Figure 8. Response accuracy (% errors) and response time as a function of the structure of the monosyllabic counterpart and the number of syllables in Experiment 5 (C2 detection). Error bars reflect 95% confidence intervals constructed for the difference between the means of sonority rises and falls.

3.3.2 Results and discussion.   Mean response accuracy (proportion errors) and response time (ms) for the correct detection of the second onset consonants are presented in Figure 8. An inspection of the means suggests that performance in this task was highly accurate and rapid. Crucially, however, ill-formed monosyllables of falling sonority (e.g., *mdif*) did not selectively impair the detection of the consonant target.

A 2 syllable × 2 onset-type ANOVA on response time yielded significant main effects of syllable, $F(1, 17) = 33.62$, $MSE = 1915$, $p < .0003$, and onset type, $F(1, 17) = 30.34$, $MSE = 1449$, $p < .0004$, and no evidence for an interaction ($F < 1$). Thus, participants took longer to detect the target in disyllables compared to monosyllables. Similarly, people took longer to detect the target *d*—the second consonant in sonority falls—compared to the target *l*—the second consonant in sonority rises. Crucially, there was no evidence that the detection of the target consonant was modulated by the grammatical well-formedness of the input. In particular, the target *d* was just as difficult to detect in the presence of ill-formed monosyllables *mdif* as it was in the presence of their well-formed disyllabic counterparts. Since the difficulty in the detection of *d* occurs across the board—irrespective of well-formedness—this effect cannot result from phonological ill-formedness of sonority falls per se.

Similar conclusions emerged from the analysis of response accuracy. Here, the ANOVA did yield a reliable interaction, $F(1, 17) = 4.68$, $MSE = .001$, $p < .05$, but Tukey HSD tests made it clear that this effect is strictly due to disyllabic counterparts. Specifically, responses to monosyllabic onsets of falling sonority did not differ reliably from those to onsets of rising sonority ($p = .99$, n.s.). The sonority of the monosyllabic onset, however, did modulate responses to the disyllabic counterparts: participants committed more errors in detecting the target *d* (e.g., for *mədif*) compared to the target *l* (e.g., in *məlif*, $p < .02$).[2]

Why is *d* harder to detect? Because this effect is not specific to monosyllables—the *mdif-mlif* contrast is either non-significant (in response accuracy) or comparable to the *medif-melif* contrast (in response time)—this phenomenon must be unrelated to sonority profile per se. Instead, the main effects of onset type, as well as that of syllable, have a simple phonetic explanation. Compared to the target *d*, the consonant *l* is marked by stronger phonetic cues—both internal cues and

anticipatory cues that are made salient by the schwa (Heid & Hawkins, 2000; West, 1999). The sensitivity of participants to such cues confirms that the segment-detection task effectively promotes attention to phonetic detail.

4 General discussion

Are ill-formed structures misidentified because of active recoding or passive failure to perceive their phonetic form? To address this question, we examined whether structures that are universally ill-formed are invariably harder to identify than better-formed structures that are unattested in the participants' native language. Five experiments compared the identification of ill-formed onsets of falling sonority and their better-formed rising-sonority counterparts under two sets of tasks. In one set, participants were asked to identify the number of syllables—a procedure that calls attention to global phonological structure; a second set of tasks elicited attention to local phonetic detail by promoting the detection of specific segments—either the pretonic schwa or the second consonant. Results showed that the two sets of procedures markedly diverged on their outcomes. When people attended to global phonological structure, ill-formed onsets of falling sonority were misidentified as disyllabic. In contrast, once their attention was focused on local phonetic detail, sonority falls were not harder to identify than their better-formed counterparts. Specifically, monosyllables of falling sonority (e.g., *mdif*) produced performance comparable to sonority rises (e.g., *mlif*) in the schwa-detection task (in Experiments 2 and 4), and sonority falls likewise presented no selective difficulty for consonant detection (in Experiment 5). Moreover, the detection of local phonetic detail was coupled with greater attention to non-distinctive phonetic cues, either cues associated with splicing of the schwa (in Experiment 2) or the salience of the second consonant (in Experiment 5). These results confirm that the two sets of tasks indeed promoted different kinds of processing. Crucially, once people attended to local phonetic detail, the disadvantage of ill-formed onsets was no longer evident.

The systematic, persistent divergence between the outcomes of the two sets of tasks cannot be blamed on artifacts that are specific to either set. It is unlikely that the misidentifications of the sonority fall endpoint in the syllable-counting tasks are artifacts of meta-linguistic knowledge: although linguistic awareness of syllables could inform syllable counting, it cannot explain the different outcomes with small sonority rises and falls (structures that are both unattested in English), and the outcomes of syllable counting converge with previous results from various other tasks (e.g., identity judgment, lexical decision, Berent et al., 2007). It is also unlikely that the elimination of ill-formedness effects in the phonetic tasks—the schwa- and consonant-detection tasks—is due to their inherent insensitivity, as each such case produced ample sensitivity to fine-grained phonetic detail. Note that we do not reject the possibility that the dispreference of ill-formed onsets is ultimately grounded in their phonetic properties, which might present a perceptual challenge. For this reason, we do not rule out the possibility that some future phonetic manipulations might be able to detect greater difficulty in the processing of ill-formed onsets. But inasmuch as our four phonetic manipulations have systematically failed to detect such difficulties, whereas tasks calling attention to global phonological structure yielded robust effects of misidentifications given identical stimuli, identical number of trials and sample sizes, it is unlikely that the phonetic reasons alone are sufficient to explain the misidentification of ill-formed structures. This suggests that misidentifications result from speakers actively recoding this (intact) phonetic form, based on their phonological knowledge.

How might phonological knowledge lead to the recoding of ill-formed onsets? According to Optimality Theory (Prince & Smolensky, 1993/2004), the phonological grammar optimizes phonological representations relative to two sets of universal grammatical constraints. One set bans ill-formed structures; another set assures the faithfulness of linguistic representations to the input. The

representation computed by the grammar depends on the ranking of these two sets of constraints. If the well-formedness constraints against a given structure (e.g., against sonority falls) are ranked above the faithfulness constraints that ensure accurate encoding, then the representation of inputs violating these constraints will typically not be faithful. Instead, it will be recoded as some better-formed output. Because the English grammar disallows any type of nasal-initial onset cluster, inputs such as *mlif* or *mdif* are both unlikely to emerge faithfully. Assuming, however, that faithfulness constraints can be promoted with uniform probability (Anttila, 1997; Davidson, Jusczyk, & Smolensky, 2006), then the probability of obtaining a faithful output will depend on the ill-formedness of the input—the worse-formed the input, the less likely the output to emerge faithfully. Thus, ill-formed onsets of falling sonority are less likely to be faithfully encoded by the grammar, and instead, they will be recoded as better-formed structures (like *medif*; for a detailed account, see Berent et al., 2009).

Although it is conceivable that people might recode speech inputs based on knowledge that is phonetic, rather than phonological, several aspects of our results specifically favor a phonological locus of repair. First, the finding that heightened attention to surface form promotes sensitivity to phonetic cues, including coarticulatory information, indicates that the (faithful) form consulted by participants is phonetic. Accordingly, active recoding must have taken place at a (higher) phonological level. Moreover, other results have documented an effect of ill-formedness even when phonetic recoding is unlikely. For example, people experience difficulties in processing onsets of falling sonority even with printed (visual) inputs (Berent & Lennertz, 2010; Berent et al., 2009). Likewise, the aversion to sonority falls taints processing their disyllabic counterparts (Berent et al., 2007, 2008)—forms that are not expected to pose any particular phonetic challenges. Such results suggest that misidentification of ill-formed structures and their aversion occurs at the phonological stage.

These conclusions do not exclude the possibility that generic properties of the auditory and articulatory interfaces might shape the grammatical phonological system, nor do they rule out the occasional occurrence of passive misperception for strictly auditory reasons. Our results, however, demonstrate that misidentification might also originate from grammatical phonological recoding. Inasmuch as phonological knowledge actively shapes identification, the relationship between the phonological grammar and misidentification must be bidirectional.

Acknowledgements

Notes

1   The onset of the pretonic vowel was defined by the increase in higher frequencies and the visible change in the amplitude and periodicity of the waveform; its offset was marked by either the onset of the stop-gap (before stops) or the small drop in frequency of the first and second formants (before liquids).
2   Similar analyses, performed on trials in which the target (*d* or *l*) was absent found no evidence that performance was modulated by well-formedness. Specifically, a 2 syllable × 2 onset type ANOVA on response accuracy found no significant effects (all $p > .16$); in response time, there was only a main effect of syllable, $F(1, 17) = 62.33$, $MSE = 1360$, $p < .0001$, due to the fact that responses to monosyllables ($M = 778$ ms) were faster than to disyllables ($M = 874$ ms).

References

Anttila, A. (1997). Deriving variation from grammar. In F. Hinskens, R. van Hout, & L. Wetzels (Eds.), *Variation, change and phonological theory* (pp. 35–68). Amsterdam: John Benjamins.

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. *Language*, *87*, 84–125.

Berent, I., Balaban, E., Lennertz, T., & Vaknin-Nusbaum, V. (2010). Phonological universals constrain the processing of nonspeech. *Journal of Experimental Psychology: General*, *139*, 418–435.

Berent, I., & Lennertz, T. (2010). Universal constraints on the sound structure of language: Phonological or acoustic? *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 212–223.

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*, 5321–5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology*, *26*, 75–108.

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.

Blevins, J. (2004). *Evolutionary phonology*. Cambridge: Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*, 117–165.

Blevins, J. (2007). Interpreting misperception: Beauty is in the eye of the beholder. In M. J. Sole, P. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 144–154). Oxford: Oxford University Press.

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: Cambridge University Press.

Davidson, L. (2006). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science*, *30*, 837–862.

Davidson, L., Jusczyk, P., & Smolensky, P. (2006). Optimality in language acquisition I: The initial and final state of the phonological grammar. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to Optimality-Theoretic grammar* (pp. 231–278). Cambridge, MA: MIT Press.

de Lacy, P. (2006). Transmissibility and the role of the phonological component. *Theoretical Linguistics*, *32*, 185–196.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*(3), 199–210.

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, CA: Stanford University Press.

Hayes, B., & Steriade, D. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 1–33). Cambridge: Cambridge University Press.

Heid, S., & Hawkins, S. (2000). *An acoustical study of long domain /r/ and /l/ coarticulation*. Paper presented at the Proceedings of the 5th Seminar on Speech Production: Models and Data & Crest Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling, Munich.

Hyman, L. (2008). Universals in phonology. *The Linguistic Review*, *25*, 83–137.

Jakobson, R. (1968). *Child language aphasia and phonological universals*. The Hague: Mouton.

Moore, B. J. C. (2003). *An introduction to the psychology of hearing*, 5th edn. London: Elsevier.

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*, 55–71.

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

Ohala, J. J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Regional Meetings, Chicago Linguistic Society*, *2*, 319–338.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, *36*, 55–90.

Peperkamp, S., & Dupoux, E. (2003). Reinterpreting loanword adaptations: The role of perception. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 367–370). Sommerville, MA: Cascadilla Press.

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Smolensky, P. (2006). Optimality in phonology II: Markedness, feature domains, and local constraint conjunction. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to Optimality-Theoretic grammar* (*Vol. 2. Linguistic and Philosophical Implications*, pp. 27–160). Cambridge, MA: MIT Press.

West, P. (1999). Perception of distributed coarticulatory properties of English /l/ and /r. *Journal of Phonetics*, *27*, 405–426.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.

Wright, R. (2004). A review of perceptual cues and robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge: Cambridge University Press.

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language*, *83*, 277–316.

Appendix. The experimental monosyllabic items used in Experiments 3–6.

Sonority rise	Sonority fall
mlɪf	mdɪf
mlɛf	mdɛf
mlæk	mdæk
mlɛb	mdɛb
mlʌp	mdʌp
mlɛk	mdɛk