Published in final edited form as:

Phonology. 2009 ; 26(1): 75–108. doi:10.1017/S0952675709001729.

Listeners' knowledge of phonological universals: Evidence from nasal clusters

Iris Berent,
Northeastern University

Tracy Lennertz,
Northeastern University

Paul Smolensky, and
Johns Hopkins University

Vered Vaknin
University of Haifa

Abstract

Optimality Theory explains typological markedness implications by proposing that all speakers possess universal constraints penalizing marked structure, irrespective of the evidence provided by their language (Prince & Smolensky, 1993/2004). An account of phonological perception sketched here entails that markedness constraints reveal their presence by inducing perceptual 'repairs' to structures ungrammatical in the hearer's language. As onset clusters of falling sonority are typologically marked relative to those of rising sonority (Greenberg, 1978), we examine English speakers' perception of nasal-initial clusters—lacking in English. We find greater accuracy for rising-sonority clusters, evidencing knowledge of markedness constraints favoring such onset clusters. The misperception of sonority falls cannot be accounted for by stimulus artifacts (the materials are perceived accurately by speakers of Russian—a language allowing nasal-initial clusters) nor by phonetic failure (English speakers misperceive falls even with printed materials) nor by putative relations of such onsets to the statistics of the English lexicon.

1. Preliminaries

1.1 Overview

It is a central tenet of Optimality Theory (OT, Prince & Smolensky 1993/2004) that robust cross-linguistic markedness generalizations arise because speakers of all languages share a system of well-formedness constraints that includes constraints penalizing marked structures. In the core case, an implicational universal of the form "Any language that admits value M on structural dimension *d* also admits value U on dimension *d*" is claimed to be a consequence of markedness constraints that assign higher well-formedness penalty to M than to U, present in all speakers' grammars: a language *L* that admits M must do so via faithfulness constraints preserving structure on dimension *d* which in the grammar of *L* out-rank the markedness constraints violated by M; these faithfulness constraints must then, by transitivity of constraint domination, out-rank the constraints violated by U, admitting U into the language as well (Prince & Smolensky 1993/2004: Ch. 9). In the case relevant to the present study, *d* is the dimension of sonority cline in onset clusters: M, the marked value, is decreasing sonority; U, the unmarked value, is increasing sonority (Section 1.2). In its most

Address for correspondence: Iris Berent, Department of Psychology, Northeastern University, 125 Nightingale, 360 Huntington Ave, Boston MA 02115, Phone (617) 373-4033, i.berent@neu.edu.

straightforward interpretation, the OT explanation of universals entails that even speakers of a language banning both M and U must possess markedness constraints that assess higher penalty to M than to U—despite the lack of direct evidence pertaining to the M/U distinction. It is this prediction that we seek to test in the present work.

How can we evaluate whether a speaker of a language possesses such knowledge of relative markedness? Like most grammatical theories, OT is fundamentally formulated to specify a 'production' rather than a 'perception' function: given an underlying form as input, an OT grammar determines the corresponding output, a surface form. It is therefore most direct to assess grammatical knowledge by examining production: one asks whether speakers of a language will reveal, in production, their knowledge that M is more marked than U, even when both are absent from their language (Section 1.3). Of course, under the most straightforward interpretation of OT, a speaker of such a language should simply fail to produce both U and M. Stochastic formulations of OT can, however, yield the prediction that production of U will sometimes succeed, and more frequently than production of M, revealing the speaker's knowledge of the markedness of M relative to U (Section 1.4.4).

Since failure to produce an ungrammatical form may be due to factors other than grammar (say, simple lack of practice executing the required articulatory programs), it is useful to complement such production-directed studies with perception-directed studies. As in production, in perception, failure to correctly process an ungrammatical form may be due to factors other than grammar (say, simple lack of practice processing the required acoustic cues)—but these factors are arguably less severe than in production and at the very least different from them, so that particular limitations of production studies may potentially be overcome by perception studies, and vice versa.

That ungrammaticality reduces the predicted accuracy of perception as well as production follows from an extension of OT to phonological perception which we sketch in Section 1.4 below: it is a form of analysis-by-synthesis (i.e., a generative model of perception), and as such preserves much of the fundamental 'synthesis' orientation of OT. Previous experimental work suggests the empirical soundness of the proposition that ungrammaticality reduces perceptual accuracy (Section 1.3), and lays the groundwork for the perception experiments that form the primary contributions of this article (Sections 2–3).

These experiments, we argue, provide new evidence supporting the premise of OT that speakers possess systems of markedness constraints that distinguish the degree of markedness of structures U and M even when neither is present in their language. Specifically, U is instantiated as the rising-sonority nasal-initial heterorganic onset clusters *ml* and *nw*, and M as the typologically more-marked falling-sonority nasal-initial heterorganic onset clusters *md* and *nb*. The case of primary interest is English-speaking participants, for whom all nasal-initial clusters are absent from their language; a comparison group is Russian-speaking participants (Sections 3.1–3.2), whose language includes both sonority-increasing and sonority-decreasing nasal-initial onset clusters.

That the knowledge distinguishing the U from the M clusters takes the form of general constraints on sonority sequencing is consistent with previous perceptual experiments examining a range of other cluster types in speakers of English, Russian and Korean (Section 1.3). We do not, however, make any particular claim about the exact form of the knowledge which informs speakers that the M clusters are more marked than the U clusters.

Whether English speakers can acquire from their linguistic experience the knowledge that the falling-sonority M onsets are marked relative to the rising-sonority U onsets is a further question about which we make no particular claim. At the segment level, we cannot exclude the possibility that learners form empirical generalizations such as 'stops are not second elements in onset clusters except following *s*', which specifically targets our falling-sonority clusters. The present results do, however, add to the challenge of precisely formulating a falsifiable theory of phonological learning, contributing further evidence that the learner must end up formulating just those generalizations that coincide with sonority sequencing principles and not others that contradict those principles. Several explicit proposals in the literature for accounting for perceptual accuracy via segment- and feature-co-occurrence statistics of the English lexicon can be examined: through a series of regression analyses, we show that they fail to account for our experimental results (Sections 4.1–4.2).

Our results do show that the relevant knowledge resides at the phonological level, rather than at lower phonetic or acoustic levels, as the preference for U to M emerges irrespective of the modality of the stimuli, for both auditory and printed words (Section 3.2).

1.2 Onset cluster markedness and sonority sequencing

A number of universals concerning the sequencing of consonants in word-initial and word-final clusters (Greenberg, 1978) can be subsumed under the principle that in unmarked cases the sonority of these consonants increase in initial and decrease in final clusters (e.g., Kiparsky, 1979; Steriade, 1982; Selkirk, 1984; Clements, 1990; Parker, 2002; Zec, 2007; see also Saussure, 1915/1959; Vennemann, 1972; Hooper, 1976). Examples for initial clusters include Greenberg's universals 17—languages admitting a (falling-sonority) liquid-obstruent sequence as in *lba* also admit a (rising-sonority) obstruent-liquid sequence as in *bla*—and 24—the presence in a language of (falling-sonority) liquid-nasal sequences such as *lma* entails the presence in that language of (rising-sonority) nasal-liquid sequences like *mla*.

Whether sonority provides the best explanation for such universals has been disputed on the basis of the phoneticians' failure to identify clear measurable correlates of sonority as well as the alleged circularity of sonority-based argumentation (e.g., Ohala, 1990). It has also been proposed that explanations be based directly on the acoustic and articulatory covariates of sonority rather than on sonority per se (Ohala, 1990; Kawasaki-Fukumori, 1992; Wright, 2004; Oudeyer, 2005). Nonetheless, the explanatory value of sonority has frequently been defended in accounts of syllable structure (Vennemann, 1972; Hooper, 1976; Steriade, 1982; Selkirk, 1984; Prince & Smolensky, 1993/2004; Smolensky, 2006); syllable contact (Vennemann, 1972, Gouskova, 2001; Gouskova, 2004); stress assignment (de Lacy, 2007) reduplication (Pinker & Birdsong, 1979; Steriade, 1982; Steriade, 1988; Morelli, 1999; Parker, 2002) and the choice of repair strategy for marked structures (Hooper, 1976).

Sonority has also proved explanatorily valuable in accounts of external evidence. The sonority of consonants correlates with their production accuracy in first- (Pater, 2004; Barlow, 2005) and second-language acquisition (Broselow & Finer, 1991; Broselow *et al.*, 1998; Broselow & Xu, 2004), developmental phonological disorders (e.g., Gierut, 1999; Barlow, 2001), aphasia (e.g., Romani & Calabrese, 1998; Stenneken *et al.*, 2005), speech errors (Stemberger & Treiman, 1986), word games (Treiman, 1984; Treiman & Danis, 1988; Fowler *et al.*, 1993; Treiman *et al.*, 2002) and reading tasks (Levitt *et al.*, 1991; Alonzo & Taft, 2002). As discussed in Section 1.3, sonority sequencing has also accounted for phonological perception data of direct relevance to the present work.

For the purposes of the experiments reported here, what is crucial is that sonority sequencing entails the markedness of the nasal-stop sequences *md* and *nb* relative to the nasal-approximant sequences *ml* and *nw*. Any comprehensive account of onset cluster sequencing entailing this markedness relation (and all those figuring in the evidence referred to above) would serve our immediate purposes equally well.

Turning from the question of the particular form of speakers' knowledge of consonant sequencing to the question of the origin of this knowledge, it is notable that most of the existing evidence involves knowledge that could in principle be projected rather directly from linguistic experience because the clusters involved are present in the speaker's language. The present studies instead examine clusters unattested in the speaker's language; the systematic behavior we observe is therefore dependent on the speaker's knowledge of general principles that extend considerably beyond direct experience.

1.3 Production and perception of unattested onset clusters

Only a handful of previous studies have examined the markedness of unattested onsets. The results suggest that unattested onsets that are relatively marked are judged as less frequent (Pertz & Bever, 1975), are less accurately produced (Broselow & Finer, 1991, Davidson, 2000; Davidson, 2006, Davidson *et al.*, 2006) and are less accurately perceived (Moreton, 2002).

Most relevant to the present work is a recent set of experiments by Berent and colleagues (2007) investigating English speakers' perception of a range of unattested 2-consonant onset clusters containing stops, including clusters with rising and falling sonority (e.g., *bn* vs. *lb*). These experiments exploited the well-known finding that speakers often misperceive[1] phonological structures that are unattested in their language (e.g., Massaro & Cohen, 1983; Hallé *et al.*, 1998; Dupoux *et al.*, 1999; Dupoux *et al.*, 2001): an utterance, produced by a speaker for whom the target form is well-formed, is perceived as having a different form by a hearer for whom the target form is ungrammatical. Thus *ebzo* as produced by a French speaker is often perceived as *ebuzo* by Japanese speakers (Dupoux *et al.*, 1999). As in this example, such misperceptions tend to "repair" the unattested form, rendering it well-formed in the hearer's language. When the unattested structure is an onset cluster, the repair often takes the form of inserting a vowel between the consonants (e.g., *tla* → *təla*; Pitt, 1998): perceptual epenthesis.

Berent *et al.* (2007) find that the likelihood of perceptual epenthesis depends on the grammatical markedness of the onset: falling-sonority onsets (e.g., *lba*) are more likely to elicit misperception (e.g., *lba* → *ləba*) than rising-sonority onsets (e.g., *bna* → *bəna*). The use of a large number of matched items allowed Berent and colleagues to demonstrate the statistical reliability of the effect of sonority cline across items. Additional experiments suggested that the misperception of marked onsets is not simply due to listeners' inability to detect the relevant phonetic cues (unattested onsets are perceived accurately given conditions that encourage closer attention to their phonetic properties; see their Experiments 5–6). Furthermore, additional analyses showed that the effect cannot be explained by several statistical properties of the English lexicon putatively relevant to perceptual accuracy (see also Albright, 2007). Indeed, subsequent research has replicated the contrast between rising- and falling-sonority onset clusters among speakers of Korean—whose lexicon arguably lacks onset clusters altogether (Berent *et al.*, 2008).

Thus, speakers of English and Korean systematically differentiate among onset clusters unattested in their language, treating those that are typologically more marked—with falling sonority—as more ill-formed, or dispreferred, in the sense of having a greater likelihood of

[1]As elaborated in Section 1.4.2, we use the term 'misperception' to refer to the computation of an unfaithful representation for nonnative inputs (relative to a native speaker's representation of those inputs). Our use of the term is neutral regarding the locus of the unfaithfulness—whether it involves inaccurate encoding of the phonetic form of the utterance or an unfaithful encoding at the phonological level of surface form resulting from grammatical constraints. Thus, our use of the term differs from some of the existing literature (e.g., on loanword adaptation), which identifies misperception specifically with a failure to extract the phonetic form of the utterance (e.g., Silverman, 1992; Peperkamp & Dupoux, 2003; Yip, 2006).

"repair" by a kind of perceptual epenthesis. An empirical limitation of these studies, however, is that they involve only clusters containing stops. For English speakers, at least, a preference for stop-sonorant sequences (a general type attested in English) over sonorant-stop sequences (unattested in English) would suffice to account for the results.[2] If speakers' knowledge truly pertains to sonority sequencing more generally, however, it should apply to other types of sequences further removed from the English inventory; in particular, among nasal-initial onset clusters—entirely lacking in English—such general knowledge would favor increasing over decreasing sonority.

The new experiments reported here test this specific prediction. To isolate the markedness of sonority sequencing from markedness related to the Obligatory Contour Principle (OCP; Leben, 1973; Goldsmith, 1976; McCarthy, 1979), we examine here only sequences obeying the OCP for manner and place. The former restriction eliminates nasal-nasal sequences (manifesting a sonority plateau that would be expected, absent OCP effects, to be intermediate in markedness between rising- and falling-sonority clusters). Because of the prohibition of ŋ from onsets in English, another constraint not under study here, the nasal consonants were restricted to *n* and *m*. The OCP-place respecting sequences we studied were *ml/nw* (rising) and *md/nb* (falling). To a first approximation, these are equated for place-markedness in the sense that all involve one labial and one coronal; they also respect the markedness constraint against the sequence voiced-voiceless in an onset (Greenberg, 1978).

On the basis of the results of Berent *et al.*, (2007), we take the hypothesis that English speakers possess general knowledge of the markedness of falling- relative to rising-sonority onsets to make the prediction of a greater likelihood of perceptual epenthesis for the falling-sonority onsets. To complement this empirically-driven prediction of greater accuracy for less marked forms, in the next section we offer a brief sketch of an OT-based account of phonological perception that provides a theoretical basis for such a prediction.

1.4. Phonological perception in Optimality Theory

The extension of standard OT that we sketch here is a modest one, intended only to link the grammaticality of phonological structures to the accuracy of their perception. Space limitations necessitate a number of omissions and simplifications. For more extensive proposals, the reader is referred to Smolensky (1996); Tesar (1997; 1998; 1999); Boersma (1998; 2007); Pater (2004); Moreton (2007); and Boersma & Hamann (2008).

1.4.1 Representations and knowledge—For our purposes it is useful to consider four levels of representation: the standard underlying form /uf/ and surface form [sf] of generative phonology, plus two additional levels. Whereas /uf/ and [sf] are discrete representations, the other two levels are continuous. One of these we'll call "phonetic form" |φf|, understood here as an encoding of acoustic-phonetic and articulatory-phonetic structure. The other continuous level we'll call "auditory form" {af}, a low-level auditory encoding of speech, perhaps something like a cochleogram. The 'input' to the speech perception system will be taken to be {af}, the external stimulus as pre-processed by the auditory system.

A full linguistic representation then will be a 4-tuple (/uf/, [sf], |φf|, {af}); whenever we refer to a 'candidate' representation, we will always mean such a 4-tuple. Each consecutive

[2]Participants in Berent et al.'s (2007) experiments also exhibited a preference for obstruent-sonorant onsets with small sonority rises (e.g., *bn*) over obstruent-obstruent onsets of level sonority (e.g., *bd*), which, in turn, were preferred to sonorant-obstruent onsets of falling sonority (e.g., *bd*). Although these preferences are consistent with the hypothesis that small sonority differences are universally marked, these results, too, could be captured by language-particular preferences. Specifically, the *bn>lb* preference might reflect a preference for onsets that begin with an obstruent, and the *bn>bd* preference might reflect a preference for a sonorant in the second position of the onset. We return to discuss these explanations in Section 4.

pair of representations constitutes a *component* of the overall system (see Fig. 1). The (/uf/, [sf]) component we take to be standard OT phonology, encoding grammatical knowledge: a ranking of Markedness constraints $M_{[sf]}$ evaluating [sf] and Faithfulness constraints $F_{/uf/,[sf]}$ evaluating the identity of /uf/ and [sf]. For the remaining components, for our purposes we assume knowledge to take a form roughly corresponding to Faithfulness constraints (for similar proposals, see Pater, 2004;Boersma & Hamann, to appear). The knowledge in the ([sf], |φf|) component will be taken to correspond to constraints $F_{[sf],|\varphi f|}$ that evaluate whether the phonetic structure of |φf| matches the speaker's realization of [sf]. Similarly, in the remaining component, knowledge corresponds to constraints $F_{|\varphi f|,|af|}$ evaluating whether the auditory structure of {af} matches the phonetic structure |φf|; this includes language-specific knowledge of the acoustic correlates of phonetic features as realized in the hearer's language.

For a given surface form such as [mlɪf], the corresponding phonetic form for the speaker's language *L* will be denoted $|\text{mlɪf}|_L$, although it should be kept in mind that like all our phonetic forms this is a continuous representation, not one with discrete segmental structure. The pair ([mlɪf], $|\text{mlɪf}|_L$) incurs no violation of the Faithfulness constraints $F_{[sf],|\varphi f|}$. Similarly, given this phonetic form $|\text{mlɪf}|_L$, there is a continuous auditory representation which we denote $\{\text{mlɪf}\}_L$ such that the pair ($|\text{mlɪf}|_L$, $\{\text{mlɪf}\}_L$) satisfies $F_{|\varphi f|,\{af\}}$. Henceforth *L* will not be notated explicitly.

The different cognitive tasks we deploy in our experiments differentially depend on various levels of representation. Our syllable-count task ('does mlɪf consist of one syllable or two?') presumably depends on the level at which syllables are encoded: [sf]. Our transcription task requires the information present at [sf] as well. An AX identity judgment ('is mlɪf identical to məlɪf?') can in principle tap any level of representation; task parameters—such as the time interval between presentation of the two forms to be judged—will modulate the relative magnitudes of the contributions of the various levels. Long inter-stimulus time intervals will favor those levels for which non-immediate memory is most robust, decreasing the role of {af} and 0259φf0259 and increasing the role of [sf] (see Section 2.3)3.

What we are proposing is a generative model of perception ('analysis-by-synthesis'): the computation that the perceptual system must perform is to find the best multi-level representation of a /uf/ that generates a [sf] that generates a 0259φf0259 that generates the input{af}. We next make precise what we mean by 'best'.

1.4.2 Evaluation—The evaluation of linguistic representations by the constraints embodying the knowledge in the three components can be recorded in the form of a kind of highly schematic tableau, as in (1). The only constraint violations of candidate A = (/mlɪf/, [mlɪf], |mlɪf|, {mlɪf}) are violations of Markedness by [mlɪf]: it is a *globally faithful* candidate in that it satisfies the Faithfulness constraints in all three components.

(1)

	/uf/	$F_{/uf/,[sf]}$	$M_{[sf]}$	[sf]	$F_{[sf],\lvert\varphi f\rvert}$	|φf|	$F_{\lvert\varphi f\rvert,\lvert af\rvert}$	{af}
A	/mlɪf/		*	[mlɪf]		|mlɪf|		{mlɪf}

Candidates that are not globally faithful are shown in tableau (2), which adopts a number of abbreviations. The horizontal location of an 'F' between columns for two levels of

[3]While it may or may not play a major direct role in our non-lexical experimental tasks, underlying form plays an indispensable role in this account because it is crucial for distinguishing grammatical from ungrammatical surface forms: the former, but not the latter, are surface forms which are optimal for some underlying form /uf/. This depends on the relative ranking of Markedness and Faithfulness constraints in the grammar, and faithfulness constraints demand an underlying form.

representation conveys that it denotes the Faithfulness constraints between those particular levels, so we omit the subscripts of M and F. The heading of each column shows the level of representation of each form so we omit from the forms the redundant delimiters distinguishing the levels. Since we will only be interested in the onset clusters here, we write only that part of each form. As before, candidate A is globally faithful. B is a candidate with epenthesis in [sf], incurring a violation of Faithfulness between /uf/ and [sf]. B has phonetic form |ml|, which is Faithful to [ml] not to [məl], so B also incurs a violation of Faithfulness between [sf] and |φf|. Candidates C and D each incur Faithfulness violations, but in different components.

(2)

	/uf/	F	M	[sf]	F	|φf|	F	{af}
A	Ml		*	ml		ml		ml
B	Ml	*		məl	*	ml		ml
C	məl			məl	*	ml		ml
D	məl			məl		məl	*	ml

In our model of perception, the auditory representation {af} plays the role of the 'input'. In that context, candidates B–C exhibit 'phonological perceptual epenthesis': the globally faithful candidate A has [sf] = [ml] (the auditory representation is {ml}) but candidates B–C have [sf]=[məl] and the faithful phonetic form |ml|. Candidate D exhibits 'phonetic perceptual epenthesis': the phonetic form is |məl|, which would be faithful to the auditory input {məl} but is unfaithful to D's actual auditory form {ml}.

We presume a continuum of degrees of Faithfulness violation in components involving continuous representations but for our very limited purposes here we can simply denote nonzero violation by a '*' in tableaux. When a candidate X violates $F_{[sf],|\varphi f|}$ to a lesser degree than candidate Y, we write X $\succ_{[sf],|\varphi f|}$ Y, with $\succ_{|\varphi f|,|af|}$ defined analogously. X $\succ_{/uf/,[sf]}$ Y means that in the (/uf/,[sf]) component, either X is optimal and Y is not, or that both are suboptimal and X has higher Harmony than Y, as standardly defined in OT.

1.4.3 Perception—We adopt a conservative approach to component interaction and assume nothing about the relative importance of constraint satisfaction in the three components. The intuition is that a possible percept can be suboptimal in one component, but only if that is required to make it optimal (or less sub-optimal) in another. This is just like constraint violation in a standard OT grammar, except that because the components are 'unranked', it is not required that sub-optimality in one component enable greater Harmony in a 'higher-ranked' component. The key point is that when an auditory form (say, that of an unattested onset cluster) would require for a fully faithful percept a surface form that is ungrammatical—and only then—there simply is no candidate (with the given auditory form) that is globally optimal: the best candidates all have sub-optimality in some component.

To formalize this explanation, for perception, we define a partial Harmony order '$\gtrdot$' among candidates—our four-level representations—as in (3). (The order is partial in that for many pairs X, Y neither X $\gtrdot$ Y nor Y $\gtrdot$ X.)

(3) *Definition*. X *has higher (perceptual) Harmony than* Y, written X $\gtrdot$ Y, if and only if X and Y have the same auditory form and either (i) or (ii) holds

- **i.** X is optimal in every component in which Y is optimal, and there is some component in which X is optimal but Y is not

ii. X and Y are optimal in exactly the same components, and in every component k in which X is not optimal, X $\succcurlyeq_k$ Y, and in some component n, X $\succ_n$ Y,

This partial Harmony order ⊳ is the basis of our perceptual account (4).

(4) *Perceptual principle*. Let X = (/x/, [x], |x|, {x}) be a globally faithful representation. Suppose given an auditory input {x}. Then a representation Y = (/uf/, [sf], |φf|, {af}) is a possible percept for {x} if and only if

{af} = {x}, and

there is no Z such that Z ⊳ Y.

When more than one percept is possible for {x}, (4) makes no assertions about the relative probabilities of the alternatives. Nonetheless, the theory has something pertinent to say about the relation between grammaticality and perception (For the demonstration of (5) and the necessary formalities, see http://www.cogsci.jhu.edu/faculty/smolensky/BerentEtAlPhonology09Addendum1.pdf).

(5) *Proposition*. Let X = (/x/, [x], |x|, {x}) be a globally faithful representation. For the auditory input {x}, there are two possibilities.

a. If [x] is grammatical—i.e., there exists /x′/ such that (/x′/, [x]) is optimal—then the only possible percept type is

A′ = (/x′/, [x], |x|, {x})

where /x′/ is any underlying form for which [x] is optimal.

b. If [x] is not grammatical, there are three possible percept types:

A′ = (/x′/, [x], |x|, {x})

C′ = (/y′/, [y], |x|, {x})

D′ = (/y′/, [y], |y|, {x})

where

[y] is the grammatical surface form most faithful to |x|,

|y| is the phonetic form faithful to [y],

/y′/ is any underlying form for which [y] is optimal, and

/x′/ is any underlying form faithful to [x].

To apply this account to onset clusters, suppose that the auditory input is {mlɪf} which we abbreviate {ml}; the phonetic form faithful to this is |ml|, the phonetic representation of an onset cluster. According to (5), for a speaker of a language (like Russian) for which *ml* is a grammatical onset, the only possible percept type is one in which the cluster is represented as such at both surface and phonetic levels. For a speaker of a language (like English) for which *ml* is not a grammatical onset, this type of percept is also possible, but there are other possibilities, which, according to (5b), depend on which grammatical surface form is most faithful to the phonetic form of the input, |ml|. For concreteness, let us assume this to be [məl]. Then, in addition to the globally faithful percept X, a possible percept for the English speaker is (/məl/, [məl], |ml|, {ml}): although the phonetic features of the cluster are faithfully perceived (|ml|), the underlying and surface forms that are perceived have the form *məl*. The other possible percept type is (/məl/, [məl], |məl|, {ml}); the phonetic features constitute a representation |məl| that is not faithful to that auditory form {ml}, but is faithful to the perceived surface form [məl].

At this point, then, the prediction concerning an auditory cluster stimulus {ml} is that Russian speakers will perceive it as a cluster at phonological and phonetic levels (because [ml] is grammatical for them) while English speakers' perceptions will be a mixture of *ml* and *məl* at all non-auditory levels: underlying, surface, and phonetic form (see also Boersma & Hamann, to appear).

An important feature of the present account is that even when the unfaithfulness of a percept is in the lowest-level component, between the phonetic and auditory representations, this unfaithful percept is only made possible by the ungrammaticality, in the highest-level component, of the globally faithful candidate. This ungrammaticality is the source of all types of unfaithfulness, for without it, the globally faithful candidate provides the only possible percept.

1.4.4 Probabilistic grammars—So far we have assumed that the grammars of Russian and English simply declare [ml] as grammatical and ungrammatical, respectively. The insufficiency of this straightforward assumption for accounting for graded performance observed experimentally has led to the adoption of various stochastic forms of OT. For example, to account for graded production accuracy for ungrammatical clusters, Davidson *et al.* 2006 proposed that while the 'base' position of Faithfulness constraints in the English grammar is such as to render a cluster like [ml] ungrammatical, speakers can promote Faithfulness constraints to higher positions by allocating additional cognitive resources; in some of these higher positions, [ml] becomes grammatical. The differential probabilities of success at producing marked clusters can be explained by rankings of Markedness constraints in the English grammar which, while normally 'hidden' because all these constraints out-rank Faithfulness and equally block output of the correspondingly marked forms, become visible as Faithfulness is stochastically promoted during production.

The same account of graded performance through stochastic OT can be applied to perception. In our case, the relevant hidden rankings are universal ones: the sonority sequencing violations of [md] are higher-ranked than those of [ml]. In English, the base position of Faithfulness is lower than the Markedness constraints violated by both clusters, but when the relevant Faithfulness constraints are promoted to some degree, they can out-rank the lower Markedness constraints violated by [ml], rendering it grammatical. With still higher promotion of Faithfulness, both [ml] and [md] become grammatical. Regardless of the relative probabilities of the different degrees of promotion, the probability that [ml] is grammatical must exceed that of [md]: every ranking that renders [md] grammatical also renders [ml] grammatical. Thus exactly the same formal structure of OT that explains implicational typological universals can explain relative accuracy rates in performance.

Combining this stochastic OT account of the English grammar with our perceptual theory (5), we see that the probability that English speakers will act like Russian speakers and necessarily perceive a cluster as such, because it is grammatical, is greater for the input {ml} than for {md}. For those rankings of the English grammar for which {ml} is ungrammatical, (5) asserts that the globally faithful percept is one possibility, despite the violation of *M([ml]) that renders it ungrammatical. The same is true for {md}, but now the fatal violation *M([md]) is even higher-ranked; if this means that the globally faithful percept for {md} is less likely than that for {ml}, it follows that {md} is less likely to be perceived as a cluster than {ml} across *all* the stochastic grammars of English.

While the details of the stochastic perception account remain to be fully fleshed out, it is reasonable to conclude that the theory sketched here predicts a higher probability for {ml} than {md} that English speakers will perceive a cluster at surface form. We now proceed to review the experiments we used to test this prediction.

1.5. Experimental tasks and predictions

The experimental paradigms we use to test the prediction of greater accuracy for less marked onsets are as follows. Experiment 1 (Section 2.1) uses a syllable count task (e.g., Does *mdɪf* consist of one syllable or two?); Experiment 2 (Section 2.2) investigates speakers' ability to distinguish monosyllabic forms from their disyllabic counterparts in an AX task (e.g., Is *mdɪf* identical to *mədɪf*?). If sonority falls are more likely to trigger perceptual epenthesis than sonority rises, then onsets of falling sonority should be more likely to be perceived as disyllabic (in Experiment 1) and to be judged as identical to their disyllabic counterparts (in Experiment 2). To seek further evidence concerning whether sonority falls are perceived less faithfully than sonority rises, Experiment 3 (Section 2.3) examines the accuracy of participants' orthographic transcription of these onsets.

To test the possibility that the misperception of sonority falls is not due to artifacts of our materials that prevent extraction of the phonetic cues for these clusters, Experiments 4–5 (Section 3.1) use the syllable-count and AX tasks to examine whether the same items can be perceived accurately by speakers of Russian—whose language allows such cluster types. If the materials are artifact-free we predict high accuracy for these speakers. To test whether difficulties processing the relevant acoustic cues are critical for English speaker's misperceptions, Experiment 6 (Section 3.2) probes the perception of stimuli presented visually: printed materials. Skilled readers are known to engage phonological representations even when they silently process printed stimuli (Orden *et al.*, 1990; Berent & Perfetti, 1995). If, as posited by our theory, the source of English speakers' perceptual inaccuracy with the auditory form {md} is ultimately grammatical—a consequence of the correct functioning of grammatical knowledge, not incorrect functioning or inadequacy of knowledge relating phonetic form to either surface or auditory form— then phonological markedness is predicted to have effects on performance with printed materials similar to its effects with auditory stimuli.

Generally speaking, cognitive preferences are often manifest in contrasts in reaction time as well as accuracy. And indeed, Berent *et al.* (2007) and Berent *et al.* (2008) found that, in addition to accuracy, response time was often observed to correlate with markedness: more marked onsets are perceived less rapidly as well as less accurately. Thus we report response time as well as accuracy measures below.

2. Are marked onsets more frequently misperceived?

2.1 Experiment 1

Experiment 1 compares sonority rises and falls for their likelihood to undergo repair in a syllable count task.

2.1.1 Participants, materials and procedure—Twenty-six native English participants, students at Florida Atlantic University, took part in the experiment in partial fulfillment of a course requirement. The materials consisted of 12 pairs of monosyllabic nonwords and their disyllabic counterparts. Monosyllabic nonwords had an unattested onset cluster and were arranged in pairs for which the onsets manifested a rise and a fall in sonority (e.g., *mlɪf, mdɪf*; see Appendix). The disyllabic materials differed from their monosyllabic counterparts only in the presence of a schwa between the onset consonants (e.g., *mlɪf*- *məlɪf*). Experiment 1 (and all subsequent experiments) also included 12 monosyllabic onsets consisting of nasal-nasal combinations and their disyllabic counterparts. These were originally included to explore speakers' perception of onsets with a sonority plateau. However, as discussed in Section 1.4, nasal-nasal sequences are additionally marked for reasons other than sonority: they violate the OCP for manner.4 Because the theory of sonority sequencing alone cannot make predictions concerning their markedness relative to the rising- and falling-sonority sequences, they will not be discussed further here. Thus, each participant was presented with a total of 72 trials: 12 item-pairs × 3 types (sonority rises, falls and plateau-fillers) × 2 syllable (mono- vs. disyllabic).

The disyllabic nonwords were produced naturally, by a female native English speaker in the sentential context "This is X" (e.g., *this is məlɪf*, with final stress). To equate the items for length, they were produced by aligning the onset of each of the words with a metronome at a rate of 100 beats per minute. The monosyllabic nonwords were next obtained by excising the pretonic vowel from the disyllabic counterpart at the zero-crossings—a procedure designed to align the two ends of the spliced waveform in order to avoid acoustic artifacts of splicing (e.g., clicks). The beginning of the vowel was defined using waveform and spectrogram inspection by the increase in the amplitude of F1 (around 660 Hz) and F2 (around 1950 Hz). The ending of the vowel was defined as follows. When the vowel preceded a nasal, the ending was defined by the decrease in energy in F1 (low F1, around 660 Hz) and the decrease in energy of F2 (around 1800 Hz, though this pattern was not consistent across items). When the vowel preceded an approximant, the vowel ending was defined by the increase in F2 (for *l*) or decrease in F2 (for *w*). Finally, vowels preceding a stop were defined by the decrease in energy in all formants associated with the stop closure. All inspections were carried out by both eye and by ear using the waveform and spectrogram. To assure that differences in responses to the monosyllabic forms are not due to differences in the salience of their counterpart as disyllables, disyllabic forms of rising and falling sonority were equated for the duration of their pretonic vowel (for sonority rises, the mean and standard deviation were M= 170 ms; SD=16 ms; for sonority falls, M=166 ms, SD=24 ms; F<1). The duration of the monosyllabic forms were M=1005 ms (SD=116 ms) and M=1097 ms (SD=87 ms), for sonority rises and falls, respectively.

Participants were seated near a computer screen wearing headphones (Sennheiser HD201). They initiated a trial by pressing the space bar, resulting in the presentation of an auditory stimulus. Participants were asked to quickly indicate whether the stimulus contained one syllable or two using the 1 and 2 keys, respectively. To illustrate the task, participants were first given a practice session with existing English words (e.g., *polite, plight*). In Experiments 1–6, trial order was randomized. Response times are reported from the stimulus onset.

2.1.2 Results and Discussion—In this and subsequent experiments we excluded outliers (responses falling 2.5 SD above or below the grand means, less than 3% of the total responses) from the analyses of response latencies. Response accuracy was analyzed as the proportion of correct responses. The effect of onset type was evaluated using ANOVAs conducted using both participants (F1) and items (F2) as random variables. Here and henceforth, when reporting statistics, 'F1' and 'F2' refer to the ANOVA F-scores by participants and by items, respectively.

Responses to disyllabic items (see Table 1) were not reliably affected by the markedness of their monosyllabic counterpart (all F<1, for response time and accuracy). However onset

[4]An inspection of the cross-linguistic typology (Greenberg, 1978, universals 9–10) suggests that onsets and codas consisting of two nasal consonants violate a restriction against a shared manner of articulation that is particularly severe for nasals. To document this fact, we compared the distribution of coda plateaus with stops, fricatives and nasals in Greenberg's typology (Greenberg provides the relevant data only for codas, but documents similar manner-restrictions for onsets and codas). Despite their identical sonority profile, the distributions of these clusters differ reliably. Not only are nasal-plateaus (19% of the sample) less frequent than fricative-plateaus (41% of the sample) and stop-plateaus (49% of the sample), but the presence of nasal plateaus implies the presence of fricative plateaus ($\chi^2(1)= 8.78$, p<.004) and stop plateaus ($\chi^2(1)=8.70$, p<.004) even after adjusting for the frequency of each of these types in the sample.

structure did modulate responses to monosyllabic items (in response accuracy: F1(1, 25)=35.61, MSE=.014, p<.0002; F2(1, 11)=14.03, MSE=.016, p<.004; in response time: both F<1): the likelihood of misjudging a monosyllabic item as disyllabic was significantly higher for items including an onset of falling sonority.

Before proceeding to seek further evidence concerning such misperceptions, we wish to briefly comment on some differences between the present findings and the earlier results of Berent et al. (2007). In Berent et al. (2007, Experiment 1), falls were mostly misperceived as disyllabic (mean correct monosyllabic responses was 14%), whereas in the present experiment, falls were mostly encoded accurately, as monosyllabic (a mean of 71%).

We note two possible explanations for this difference. The first is phonological, and depends on the magnitude of sonority cline: Relative to the clusters used in Berent et al. (2007), which were mostly liquid-obstruent combinations, the nasal-obstruent clusters used here have a smaller fall in sonority, which may lead to a smaller effect. The second possible explanation is based on phonetic differences in the materials. Unlike the clusters in Berent et al., which were recorded naturally by a Russian speaker, our present materials were produced by splicing from recordings by a native English speaker. The familiarity with the English phonetic categories might have allowed our English participants to better identify the initial consonants as adjacent. The acoustic properties of the present stimuli are considered further in Section 3.1.

Whether the advantage of nasal-initial falls relative to liquid-initial ones is due to phonetic or phonological sources is a matter for further research. For our present purposes, however, more crucial is the convergence across the two types of materials. In both cases, sonority falls were more likely to trigger disyllabic misperception than rises.

2.2 Experiment 2

Experiment 2 directly examines whether English speakers perceive marked onsets epenthetically by eliciting identity judgments. Twenty-four native English speakers, students at Florida Atlantic University, took part in the experiment in partial fulfillment of a course requirement. The materials were the same stimuli used in Experiment 1. They were arranged in pairs. Pair members were either identical (either monosyllabic or disyllabic) or epenthetically related (*mlɪf-məlɪf*; or *məlɪf-mlɪf*, with order counter-balanced). The materials were next arranged in two lists, matched for the number of stimuli per condition (target type × identity × order) and counterbalanced, such that, within a list, each item appeared in either the identity or the nonidentity condition. Each participant was presented with both lists, with order counter-balanced across participants.

Each trial began with a fixation point (*). Participants initiated the trial by pressing the space bar, triggering the presentation of the first auditory stimulus, followed by the second (onset asynchrony=1500ms). Participants responded by pressing the 1 or 2 keys, for “identical” and “non-identical” responses, respectively. Slow responses (RT >2500 ms) received a computerized warning signal. Prior to the experiment, participants were given a short practice using English words (e.g., *plight-plight* vs. *polite-plight*).

Responses to identity trials were generally accurate and fast (M=95.2%; M=1129 ms). Our main interest concerns responses to nonidentical items (see Table 2). Participants were significantly more likely to misperceive monosyllabic items as identical to their disyllabic counterparts for items with sonority falls compared to rises (F1(1, 23)=4.38, MSE=.018, p<.05; F2(1, 11)=5.63, MSE=.007, p<.04; in response time both F<1).

2.3 Experiment 3

The greater likelihood of misperceiving onsets of falling sonority is consistent with the hypothesis that the markedness of sonority falls impedes their faithful encoding. As pointed out by Peperkamp (2007), however, the existing results do not strictly demonstrate that such onsets are represented less faithfully: they show that sonority falls are more likely to undergo *epenthesis* than rises. Although there is every reason to believe that that the rate of epenthesis reflects the rate of unfaithful encoding, a divergence is logically possible—when *non*-epenthetic misperceptions are taken into account, it is conceivable that they could reverse the conclusion, with less marked onsets being overall less faithfully encoded than more marked onsets. Accordingly, it is desirable to seek converging evidence from tasks that can identify all kinds of unfaithful encoding. To this end, Experiment 3 uses a transcription procedure. In each trial, participants are presented with an auditory word, and are asked to transcribe it using English orthography. If the representation of marked onsets is less faithful, then marked onsets should be less likely to elicit correct transcriptions.

Sixteen native English speakers, students at Florida Atlantic University, took part in the experiment in partial fulfillment of a course requirement. The materials were the twelve pairs of monosyllabic nonwords used in Experiment 1. To approximate the conditions used in our previous experiments, we presented the monosyllabic items mixed with their disyllabic counterparts, which were treated as fillers (i.e., unanalyzed). Participants were seated near the computer, wearing headphones. They initiated each trial by pressing the space bar, triggering the presentation of a single auditory stimulus. Participants were asked to transcribe the item on a piece of paper using English orthography.

Participants correctly transcribed monosyllabic forms on only 45% of the trials. Errors in the transcription of the onset included epenthesis (e.g., *mlif* → *melif*; 8.3% of the total responses), prothesis (e.g., *mlif* → *emlif*; 2.8% of the total responses), consonant substitution (e.g., *mlif* → *nlif*; 13.3% of the total responses), consonant deletion (e.g., *mlif* → *lif*; 22.13% of the total responses), and others (omissions, lexicalizations and radical changes to the input, a total of 8% of the total responses). However, the rate of each of these error types was not reliably modulated by onset structure (all p's>.05). The insensitivity of the error patterns to onset type is likely due to individual differences, prompted by the susceptibility of this off-line procedure to problem-solving strategies. For example, participants could have discerned that all onsets begin with a nasal consonant, either *m* or *n*, and consequently, determine the initial consonant by guessing (a possibility supported by the high rate of C1 substitution, e.g., *mdif* → *ndif*). Because individual participants might differ on their susceptibility to such strategies and the number of participants is small, the error variance across participants is expected to be high, reducing sensitivity even further.

Despite these limitations, the transcription task was nonetheless sensitive to onset structure. A planned comparison yielded an effect of onset that was significant by items ($t2(1,11)=3.17$, $MSE=13.5$, $p<.007$) and marginally so by participants ($t1(1, 15)=1.71$, $MSE=3.88$, $p<.06$, one tail). Onsets with rising sonority were transcribed correctly on 50.5% of trials, but with falling sonority, only 40.6%.

3. The representational level of cluster misperception

The results of Experiments 1–3 demonstrate that more marked onsets of falling sonority are more likely to be misperceived than less marked onsets with a sonority rise: such marked onsets are more likely to be misperceived as disyllabic (Experiment 1); such misperception persists even when participants are explicitly asked to discriminate such items from their disyllabic counterpart (Experiment 2); and the reduced probability of faithful perception of marked onsets remains when all types of unfaithfulness are available for report in transcription (Experiment 3). We now turn to the question of the level of representation at which unfaithful encoding occurs. This topic has received considerable attention in the context of loanword adaptation, with some researchers emphasizing the role of phonetics (e.g., Peperkamp & Dupoux, 2003; Peperkamp *et al.*, 2008), others the role of phonology (e.g., Broselow & Finer, 1991; Silverman, 1992; Lacharite & Paradis, 2002; Yip, 2006). Recall that in our account of perception, unfaithfulness can occur at multiple levels, although the possibility of an unfaithful percept, even at low levels, is only made possible by ungrammaticality, at the highest level, of the globally faithful candidate (Section 1.4.3).

That these misperceptions of unattested clusters are indeed a consequence of the correct functioning of grammatical knowledge—as opposed to incorrect functioning or inadequacy of knowledge relating phonetic form to either surface or auditory form—is examined further in Section 3.2. But first we wish to deal with another possibility: that what we consider 'misperception' is merely a low-level consequence of stimulus artifacts. Perhaps the increased likelihood of perceptual epenthesis for sonority falls reflects our failure to fully remove the pretonic vowel when splicing them out of their disyllabic counterparts[5].

3.1 Is the misperception of sonority falls due to stimulus artifacts?

To assess the possibility that the 'clusters' in our stimuli are acoustically defective, failing to provide adequate evidence for a cluster, we turned to speakers for whom ungrammaticality of the relevant cluster types is not the contributing factor that it is for our English-speaking participants. Russian permits nasal-initial clusters of both increasing (e.g., *ml*) and decreasing (e.g., *mg*) sonority. Three of the particular clusters we used—*nw*, *md* and *nb*— are not specifically attested in Russian, but as we will see, our Russian participants have no difficulty perceiving these clusters; in our theoretical analysis of Section 1.4, we thus treat them as accidental gaps rather than the true cases of ungrammaticality we take them to be in English.

If our stimuli are defective in their cues for clusters, Russian speakers should, like English speakers, experience some difficulty in perceiving these stimuli as clusters. To assess this possibility, Experiment 4 uses the syllable-count task and Experiment 5 uses identity judgment.

3.1.1 Experiment 4—Twenty-six native Russian speakers, students at the University of Haifa, Israel, took part in this experiment. The materials and procedure are as described in Experiment 1 (Section 2.1), except for the use of Russian (instead of English) words in the practice phase (e.g., *drov*, 'log'; *darov*, 'present').

To gauge the effect of linguistic knowledge on the processing of onset clusters, we compared the performance of Russian speakers to English participants (in Experiment 1) separately, for monosyllabic and disyllabic inputs. Response accuracy is presented in Figure 2 (with error bars reflecting 95% confidence intervals for the difference between the means)[6]; response time is given in Table 3.

Monosyllabic items: The responses of Russian and English speakers to monosyllabic items were compared by means of a 2 Language × 2 Type ANOVA. The analyses of response time

[5]An anonymous reviewer notes that our splicing procedure may in effect have induced a bias *against* our hypothesis, since more cues for the pretonic vowel may have been removed from sonority falls than sonority rises.

[6]Note that these confidence intervals are constructed for the *difference* between means (i.e., the difference in response accuracy for sonority rises and falls), rather than for absolute means. Loftus and Masson (Loftus & Masson, 1994) showed that these two types of confidence intervals are related by a factor of $\sqrt{2}$. They further demonstrated that the difference between any two sample means is significant by a two-tailed t-test if and only if it exceeds the confidence interval constructed for the difference between those means (using the same alpha level).

yielded only a marginally significant effect of language ($F1(1, 50)=3.66$, $MSE=120,367$, $p<.07$; $F2(1, 11)=8.13$, $MSE=1675$, $p<.02$), suggesting that Russian speakers tended to respond more slowly than English participants, an effect that is likely due to their unfamiliarity with the phonetic categories of the English speaker who produced these recordings. Nonetheless, Russian speakers were significantly more accurate than English participants, resulting in a significant main effect of language in the analyses of response accuracy ($F1(1, 50)=30.93$, $MSE=.025$, $p<.0002$; $F2(1, 11)=39.59$, $MSE=.0045$, $p<.0001$). Crucially, the analyses of response accuracy yielded a significant interaction ($F1(1, 50)=26.79$, $MSE=.008$, $p<.002$; $F2(1, 11)=10.22$, $MSE=.012$, $p<.009$), suggesting that the effect of onset type was modulated by linguistic knowledge. We thus proceeded to test the effect of onset type separately for Russian speakers. The response accuracy of Russian speakers approached ceiling, and it was unaffected by onset type (both F's<1.1). Thus, unlike English speakers, Russian speakers were no more likely to misperceive onsets of falling sonority epenthetically relative to sonority rises.

Disyllabic items: The ANOVA (2 language × 2 onset type) comparing Russian and English participants yielded a significant interaction in the analysis of response accuracy ($F1(1, 50)=19.23$, $MSE=.006$, $p<.0001$; $F2(1, 11)=15.11$, $MSE=.005$, $p<.003$; for response time, no effect was reliable by participants and items, all p's>.07). Thus, responses to disyllabic items were modulated by language.

A separate analyses of the Russian group showed that the responses of Russian speakers to disyllabic items were significantly affected by onset types: Russian speakers responded more accurately to the counterparts of sonority falls than to the counterparts of rises ($F1(1, 25)=28.78$, $MSE=.01$, $p<.0002$; $F2(1, 11)=13.65$, $MSE=.010$, $p<.004$). This effect must be due to their linguistic knowledge, rather than stimulus artifacts, as English speakers were equally likely to perceive both types as disyllabic. Indeed, Russian speakers were overall less accurate than English speakers in responding to disyllabic items ($F1(1, 50)=3.49$, $MSE=.017$, $p<.07$; $F2(1, 11)=12.11$, $MSE=.009$, $p<.006$). The difficulty of Russian speakers in perceiving CəCVC items as disyllabic, a finding that agrees with previous research (Berent et al., 2007), might reflect the absence of a pretonic schwa in many dialects of Russian (Crosswhite, 1999)[7]. The unfamiliarity (or ungrammaticality) of such structures might have led to their confusion with monosyllabic forms. Interestingly, however, such confusions were more pronounced with the counterparts of sonority rises—a result consistent with previous results for both Russian and English speakers (Berent et al. 2007; Experiments 2 and 1, respectively). This may be a result of markedness-driven perceptual competition: since rising-sonority clusters are less marked, as an alternative to the CəCVC percept, CCVC is a stronger competitor when its onset cluster has rising sonority. Alternatively, this effect may be a consequence of the statistical structure of the Russian lexicon—our present results do not allow us to discriminate between these possibilities. Either way, the findings make it clear that our marked monosyllabic stimuli are perceptible as such, suggesting that the misperception of these items by English speakers is due to linguistic knowledge.

3.1.2 Experiment 5—Experiment 5 is the counterpart with Russian speakers of the identity-judgment task of Experiment 2 (Section 2.2).

[7]Unfortunately, we cannot ascertain the precise dialect spoken by participants. Since some dialects of Russian do exhibit an immediately pretonic schwa (Crosswhite, 1999), one might wonder whether the difficulty of Russian speakers with disyllabic forms might be due to knowledge of other languages, most notably Hebrew. However, Hebrew does not systematically reduce pretonic vowels, and our subsequent work with Hebrew participants experiments using the same materials yielded high response accuracy to disyllabic forms (M=93%). The contrast between the responses of Russian and Hebrew speakers suggests that the difficulty of Russian speakers with disyllabic forms is specifically due to their knowledge of Russian.

Twenty-four native Russian speakers, students at the University of Haifa, Israel, took part in this experiment. The materials and procedure are as described in Experiment 2, except the use of Russian words in the practice phase (e.g., *drov-drov, darov-drov*).

As expected, the responses of Russian speakers in the identity condition were generally fast (M=1315 ms) and accurate (M=95.7%). Our main interest is in the effect of linguistic knowledge on participants' ability to discriminate the monosyllabic forms from their disyllabic counterparts. To assess this issue, we compared the responses of English and Russian speakers to nonidentity items by means of a 2 Language × 2 Onset type analysis. The accuracy means are presented in Figure 3; response time means are given in Table 4.

The analyses of response accuracy yielded a significant main effect of language ($F1(1, 46)=7.87$, MSE=.041, $p<.008$; $F2(1, 11)=43.77$, MSE=.003, $p<.0001$), demonstrating that Russian speakers were overall more accurate than English speakers in discriminating among nonidentical items. As in the syllable count task, however, Russian speakers took longer time to respond ($F1(1, 46)=12.52$, MSE=62898, $p<.001$; $F2(1, 11)=594.13$, MSE=531.72, $p<.0001$), an effect we attribute to their unfamiliarity with the phonetic categories of the English speaker who produced these items. Importantly, however, linguistic knowledge modulated the effect of onset type, resulting in a significant interaction in the analysis of response accuracy ($F1(1, 46)=9.48$, MSE=.011, $p<.004$; $F2(1, 11)=19.13$, MSE=.003, $p<.002$; In response time, both $F<1$).

Recall that for English speakers, the likelihood of misperceiving sonority falls as identical to their disyllabic counterparts was greater than for sonority rises. In contrast, Russian speakers responded with greater accuracy to onsets of falling sonority relative to sonority rises, an effect significant by participants, and marginally so by items ($F1(1, 23)=6.90$, MSE=.005, $p<.02$; $F2(1, 11)=3.84$, MSE=.004, $p<.08$). Russian speakers' inaccuracy with onsets of rising sonority may well be due to their difficulty in perceiving the *disyllabic* form in the pair—a difficulty evident also in the syllable count experiment. In any event, the results make it clear that our materials of falling sonority are not universally confusable with their disyllabic counterparts.

3.2. The role of phonetic form in the misperception of sonority falls

We turn now to the hypothesis that English speakers' increased misperception of marked onset clusters is due to their inability to extract accurate phonetic representations of these clusters. Berent et al. (2007) argued against this hypothesis by showing that English speakers can perceive marked onsets accurately (as accurately as they perceive their unmarked counterparts) under conditions that encourage attention to phonetic detail (their Experiments 5–6).

The following experiment presents yet a stronger test of this hypothesis. Here, we examine whether the increased misperception of marked onsets emerges for stimuli that are devoid of any acoustic properties—for printed stimuli read silently. The materials and task—identity judgment—are the same as in Experiment 2 (Section 2.2), except that the words are presented visually, in alternating cases (e.g., *mlif-MELIF*). To encourage participants to commit the items to maintenance in phonological working memory—a process that requires the assembly of their phonological structure from print (Baddeley, 1986)—we increased the inter-stimulus interval to 2500 ms, an interval longer than that used in Experiments 2 and 5 (which used an onset asynchrony of 1500 ms). Previous research using this procedure with obstruent-sonorant combinations demonstrated its sensitivity to the phonological structure of the materials (Berent & Lennertz, 2008): As with auditory materials, participants took longer to distinguish among nonidentical items with sonority falls (e.g., *lbif-LEBIF*) than rises (e.g., *bnif-BENIF*). Moreover, participants were sensitive to the phonological similarity among the items even when their graphemic overlap was controlled. Specifically, people were quicker to discriminate *blim* from *kelim*—items that differ on two letters and two phonemes—than *clim-kelim*, which differ by two letters but only one phoneme. Thus, despite the use of printed materials, we expect participants in the present experiment to be sensitive to the phonological structure of the printed words. If the increased misperception of marked onsets reflects not a low-level failure in processing auditory input but rather the phonological markedness of these onsets, then similar results might emerge with printed materials.

3.2.1 Experiment 6—Twenty-four native English speakers, students at Florida Atlantic University, took part in the experiment either in partial fulfillment of a course requirement or for payment. The materials were printed nonwords corresponding to the items used in Experiment 2 (Section 2.2). The structure of the materials was as described in Experiment 2.

Each trial began with a fixation point (*) presented for 100 ms followed by the first nonword, presented in lower case for 500ms, a pattern mask (XXXXXX) presented for 2500ms and the second nonword, presented in upper case until participants made their response. The procedure was otherwise as in Experiment 2.

Four participants were excluded because their mean accuracy fell 1.5SD below the group's mean. Response time and accuracy to the identity trials were M=747 ms and M=89.7%, respectively. Our main interest is in the nonidentity trials (see Table 5). Readers responded more accurately to onsets of rising sonority than to sonority falls ($F1(1, 19)=6.08$, MSE=.005, $p<.03$; $F2(1, 11)=6.01$, MSE=.003, $p<.04$; in response time: both $F<1$).

4. The role of lexical knowledge

The experiments we have presented provide evidence that English speakers, for nasal-initial onset clusters, misperceive falling-sonority clusters more frequently than rising-sonority clusters, and that these results are not due simply to stimulus artifacts, nor to hearers' inability to perform the acoustic analysis necessary for accurate perception. This evidence is consistent with predictions based on the OT perceptual account sketched in Section 1.4 which attribute the effect to grammatical knowledge of the relative markedness of phonological surface forms containing such clusters. But on an alternative explanation, the difficulty with sonority falls reflects not their grammatical markedness, but rather statistical knowledge of the English lexicon. The hypothesis that hearers' preference for onset clusters obeying sonority sequencing principles only reflects statistical knowledge is challenged by recent findings replicating this preference in speakers of Korean—a language that arguably lacks onset clusters altogether (Berent *et al.*, 2008). Those findings, however, do not rule out the possibility that statistical knowledge might account for the performance of English participants in the present experiments. We next consider this possibility, at the levels of segmental (Section 4.1) and featural (Section 4.2) statistics of the English lexicon.

4.1 Segmental lexical statistics

To evaluate the possibility that the preference of items like *mlɪf* reflects only the co-occurrence of their segments in the English lexicon, we calculated several statistical measures of our materials, including indices of neighborhood structure, segment/letter co-occurrence and the properties of the first consonant. These measures are briefly summarized below—a detailed description of these measures, their calculation and a justification of their use may be found on (http://www.cogsci.jhu.edu/faculty/smolensky/BerentEtAlPhonology09Addendum2.pdf).

Neighborhood measures included the number of neighbors (the number words of obtained by adding, deleting or substituting one of a target's phonemes) and their summed frequency. Measures of segment co-occurrence were calculated at both the level of the whole word and the onset. Word measures included position-sensitive phonemeand bi-phone probability (for auditory words) and bigram count and bigram frequency (for printed words); Onset measure estimated the probability that the two onset-consonants co-occur by the position-sensitive probability of each of the two consonants. Finally, because fronting is known to affect auditory perception (Byrd, 1992; Surprenant & Goldstein, 1998), the status of the initial consonant (*m* vs. *n*) was also considered in the analysis of auditory words.

We next assessed the unique contribution of sonority profile and statistical knowledge to performance in the syllable-count (Experiment 1), identity (Experiments 2 and 6) and transcription (Experiment 3) tasks by means of multiple step-wise linear regression analyses using response accuracy to each of the 24 stimulus items (averaged across participants) as the dependent measure.

To address the unique contribution of sonority profile, we forced this factor as the last step into the regression analysis, after controlling for statistical properties, entered in the first step. Another analysis addressed the unique contribution of statistical knowledge (entered last) after controlling for the effect of sonority (entered first). To specifically isolate the effect of statistical knowledge concerning onset structure, we also assessed the effect of segment/letter co-occurrence in the onset and in the whole word in separate analyses. Thus, in each experiment, statistical properties were examined using either whole-word and neighborhood properties or onset and neighborhood properties, and the effect of these statistical properties was entered either first or last—a total of four analyses per experiment. Because our main interest is in the unique contribution of sonority profile and statistical properties, we only report the proportion of the change in variance associated with the last predictor (indicated as R^2 change, see Table 6).

The findings show that participants were sensitive to statistical structure. Statistical properties reliably captured up to 28.4% of the variance in Experiment 2, and 43.9% of the variance in the transcription task. In fact, once the statistical properties of the onset and neighborhood were controlled, the effect of sonority in the transcription task was eliminated altogether, possibly reflecting the vulnerability of this off-line task to guessing. Indeed, statistical knowledge did not subsume the effect of sonority in either of our on-line experiments (i.e., the syllable count and identity judgment procedures). Although the unique contribution of sonority profile was not reliable in all analyses, in no case was the effect of statistical properties reliable without the effect of sonority being either significant or marginally so. These results speak against the attribution of our findings only to the statistical distribution of segments in the English lexicon.

4.2. Featural lexical statistics

It is plausible that learning relevant to sonority sequencing takes place at the level of features, rather than segments. It is unknown how successfully our data could be accounted for by a sophisticated feature-based statistical learning model such as Hayes & Wilson, 2008 when trained on the English lexicon (although the study by Hayes, 2007 of a toy lexicon suggests that at least some degree of sonority-informed learning bias might be required). In what follows, we examine two predictions arising from simple facts of the distribution of major-class features in English: the first concerns glides following initial nasals and the second sonorants following initial consonants. Could a statistical model that simply tracks such featural distributions capture our results?

Although English lacks onset clusters that begin with a nasal consonant, nasal-glide sequences are nonetheless attested at the beginning of English words (e.g., *mule*, *mute*). We follow Giegerich (1992), Davis & Hammond (1995) and Buchwald (2005) in taking the palatal glide [j] to be part of a diphthong in the syllable nucleus rather than as part of the onset (the palatal glide can only precede [u], and, unlike [w], does not constrain the preceding consonant as would be expected if it were in the onset). But regardless of how such sequences are represented, it is conceivable that familiarity with nasal-glide sequences might inform our participants' preference for sonority rises.

Our materials allow us to evaluate this possibility. Recall that our sonority rises comprise nasal-liquid (e.g., *ml*) and nasal-glide (e.g., *nw*) combinations. If the preference for sonority rises reflect familiarity with nasal-glide sequences, then one would expect that (a) *nw*-sequences should be recognized more accurately than *ml*- sequences; and (b) the advantage of sonority rises over falls should be larger for *nw*- compared to *ml*-initial sequences (a prediction that might also follow from the greater sonority rise for nasal-glide than for nasal-liquid). A 2 (C1 type) × onset type (rise vs. fall) ANOVA comparing the response accuracy with these two types of items indeed yielded a significant interaction in Experiment 1 ($F2(1, 10)=12.44$, $MSE=.008$, $p<.006$) and a marginally significant interaction in Experiment 2 ($F2(1, 10)=3.82$, $MSE=.006$, $p<.08$; for Experiment 3 and 6, $p>.16$), but contrary to the prediction of the statistical account, response accuracy was actually higher for the *ml*-sequence, and the advantage of sonority rises over falls was likewise larger for the *m*-relative to the *n*-initial items (possibly, due to a fronting effect; see Byrd, 1992; Surprenant & Goldstein, 1998).

Although this result lends no support to the statistical explanation, the lack of a familiarity effect might be due to the limited experience of English speakers with nasal-glide sequences and their dubious status as onsets. Unlike the small frequency difference between the different types of nasal onsets of rising sonority—those with C2 comprising liquid vs. glides —as a whole, onsets of rising sonority are clearly more frequent than falls, since sonorants are far more frequent than obstruents at the second position of the onset. The preference for sonority rises over falls could reflect this difference in feature co-occurrence.

Although the results presented so far do not rule out this explanation, other aspects of the findings speak against this possibility. As noted in Section 2.1.1, our experiments included not only onsets with sonority rises and falls but also a group of onsets of level sonority (e.g., *mnɪf*, *nmɪf*) which were excluded from the analyses reported above because they violate the OCP for manner. Nonetheless, these items have some relevance for addressing the feature-statistical account. Unlike our onsets of falling sonority, whose second consonant is a voiced obstruent, never occurring in second position in English onsets, in onsets of level sonority, the second consonant is a nasal, and nasals occur quite frequently in this position in English (e.g., *snow, small*). If English speakers base their responses only on the statistics of features in English, then it is plausible that they should consistently favor nasal-second onsets like *mnif* and *nmif* to the voiced-obstruent-second onsets of falling sonority. As it happens, this prediction is consistent with the results of the syllable count task ($M_{plateau}$=82.1%; $\Delta_{fall\text{-}plateau}$=−11.2%, $t1(50)=3.38$, $p<.002$; $t2(22)=2.54$, $p<.02$). But the advantage of sonority plateaus over falls is not systematic. The difference was not significant in both the transcription and identity tasks involving orthography (in Experiment 3 $M_{plateau}$=36%, $\Delta_{fall\text{-}plateau}$ =4.69%; all $t<1$; in Experiment 6 $M_{plateau}$=87.6%, $\Delta_{fall\text{-}plateau}$ =−2.3%, $t(22)=1.18$; $t(138)=1.23$). Moreover, the identity task with auditory materials (Experiment 2) yielded significantly *lower* accuracy for level relative to falling sonority ($M_{plateau}$=62.3%, $\Delta_{fall\text{-}plateau}$ =10.3% $t1(46)=2.76$, $p<.009$; $t2(22)=2.33$, $p<.03$)—a finding that might be due to the strong violation of the OCP-place by nasal-nasal sequences. A full investigation of the interaction of OCP- and sonority-sequencing-markedness obviously falls beyond the scope of this research. For our present purposes, suffice it to note that participants did not consistently favor the more familiar, *mn*-type onsets relative to the less familiar *md*-type onsets. Thus, our results appear to lend no support to the claim that performance in our experiments reflects only statistical knowledge of the distribution of features in English onset clusters.

5. Summary

Psycholinguistic research typically concerns speakers' knowledge of structures attested in their language. Optimality Theory (Prince & Smolensky, 1993/2004), however, proposes that speakers' knowledge includes universal structural preferences concerning even structures unattested in their own language. According to our extension of OT to perception, these universal preferences are predicted to be manifest not just in production, but in perception as well, with dispreferred structures being less accurately perceived. The research reported here probed for such universal preferences regarding unattested structures by examining the perception of nasal-initial onsets by English speakers. Across languages, onset clusters—including nasal-initial ones—with falling sonority are marked relative to those of rising sonority (e.g., Greenberg, 1978). Our experimental findings are consistent with the prediction that such markedness relations are known by English speakers: Compared to those with rising sonority, nasal-initial onsets with falling sonority are more likely to be misperceived as disyllabic and to be misjudged as identical to their epenthetic counterparts (Experiments 1–2). Onsets with falling sonority are also reproduced less accurately than those of rising sonority in transcription (Experiment 3), suggesting that their encoding is less faithful.

The misperceptions of our materials with sonority falls are not due to stimulus artifacts: Russian speakers, whose language allows nasal-initial onsets with falling sonority, were no more likely to misperceive sonority falls than rises (Experiments 4–5). Similarly, these misperceptions are not due to a simple failure of English hearers to process the acoustic cues to falling-sonority clusters, as they emerge regardless of stimulus modality—for both auditory (Experiment 2) and printed materials (Experiment 6).

Examining several statistical properties of the English lexicon, at the segmental and featural levels, we are unable to explain our results using a number of statistical accounts of perceptual accuracy that have been previously proposed.

The experimental evidence and analysis presented here thus suggest that English speakers possess phonological knowledge of the relative markedness of onset clusters which, for the moment, seems unexplainable purely from their linguistic experience, but which is fully expected on the basis of universal grammar.

References

Albright, Adam. Natural classes are not enough: Biased generalization in novel onset clusters. Ms: MIT; 2007.

Alonzo, Angelo; Taft, Marcus. Sonority constraints on onset-rime cohesion: evidence from native and bilingual Filipino readers of English. Brain and language. 2002; 81:368–383. [PubMed: 12081406]

Baddeley, Alan D. Working memory. Oxford: Oxford University Press; 1986.

Barlow, Jessica A. The structure of /s/-sequences: Evidence from a disordered system. Journal of Child Language. 2001; 28:291–324. [PubMed: 11449941]

Barlow, Jessica A. Sonority Effects in the Production of Consonant Clusters by Spanish-Speaking Children. In: Eddington, David, editor. Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages; Somerville, MA: Cascadilla Proceedings Project; 2005. p. 1-14.

Berent I, Lennertz T, Jun Jongho, Miguel Moreno A, Smolensky P. Language universals in human brains. Proceedings of the National Academy of Sciences. 2008; 105:5321–5325.

Berent, Iris; Lennertz, Tracy. Universal constraints on the sound structure of language: Phonological or acoustic? 2008

Berent, Iris; Perfetti, Charles A. A rose is a REEZ: The two cycles model of phonology assembly in reading English. Psychological Review. 1995; 102:146–184.

Berent, Iris; Steriade, Donca; Lennertz, Tracy; Vaknin, Vered. What we know about what we have never heard: Evidence from perceptual illusions. Cognition. 2007; 104:591–630. [PubMed: 16934244]

Boersma, Paul. Ph.D. dissertation. University of Amsterdam; 1998. Functional phonology: formalizing the interaction between articulatory and perceptual drives.

Boersma, Paul. Some listener-oriented accounts of h-aspiré in French. Lingua. 2007; 117:1989–2054.

Boersma, Paul; Hamann, Silke. Introduction to 'Phonology in perception'. In: Boersma, Paul; Hamann, Silke, editors. Phonology in perception. Berlin: Mouton de Gruyter; (to appear)

Boersma, Paul; Silke, Hamann. Loanword adaptation as first-language phonological perception. Rutgers Optimality Archive 975. 2008

Broselow, Ellen; Chen, Su-I.; Wang, Chillin. The emergence of the unmarked in second language phonology. Studies in second language acquisition. 1998; 20:261–280.

Broselow, Ellen; Finer, Daniel. Parameter setting in second language phonology and syntax. Second Language Research. 1991; 7:35–59.

Broselow, Ellen; Xu, Zheng. Differential difficulty in the acquisition of second language phonology. International Journal of English Studies. 2004; 4:135–163.

Buchwald, Adam. Representing Sound Structure: Evidence from Aphasia. In: Alderete, John; Han, Chung-hye; Kochetov, Alexei, editors. Proceedings of the 24th West Coast Conference on Formal Linguistics; Somerville, MA: Cascadilla Proceedings Project; 2005. p. 79-87.

Byrd, Dani. Perceptual assimilation in consonant clusters: A gestural model. Phonetica. 1992; 49:1–24.

Clements, GN. The role of the sonority cycle in core syllabification. In: Kingston, J.; Beckman, M., editors. Papers in laboratory phonology I: Between the grammar and physics of speech. Cambridge: Cambridge University Press; 1990. p. 282-333.

Crosswhite, Katherine Margaret. Vowel reduction in Optimality Theory. PhD dissertation. Los Angeles: University of California; 1999.

Davidson, Lisa. Experimentally uncovering hidden strata in English phonology. In: Gleitman, L.; Joshi, A., editors. Proceedings of the 22nd Annual Conference of the Cognitive Science Society; Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

Davidson, Lisa. Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. Cognitive Science. 2006; 30:837–862. [PubMed: 21702839]

Davidson, Lisa; Jusczyk, Peter; Smolensky, Paul. Optimality in language acquisition I: The initial and final state of the phonological grammar. In: Smolensky, P.; Legendre, G., editors. The harmonic mind: From neural computation to Optimality-Theoretic grammar. Cambridge, MA: MIT press; 2006.

Davis, Stuart; Hammond, Michael. On the status of onglides in American English. Phonology. 1995; 12:159–182.

de Lacy, Paul. The interaction of tone, sonority, and prosodic structure. In: deLacy, Paul, editor. The Cambridge handbook of phonology. Cambridge: Cambridge University Press; 2007. p. 281-307.

Dupoux, Emmanuel; Kakehi, Kazohiko; Hirose, Yuki; Pallier, Christophe; Mehler, Jacques. Epenthetic vowels in Japanese: A perceptual illusion? Journal of Experimental Psychology: Human Perception and Performance. 1999; 25:1568–1578.

Dupoux, Emmanuel; Pallier, Christophe; Kakehi, Kazohiko; Mehler, Jacques. New evidence for prelexical phonological processing in word recognition. Language and Cognitive Processes. 2001; 5:491–505.

Fowler, Carol; Treiman, Rebecca; Gross, Jennifer. The structure of English syllables and polysyllables. Journal of Memory and Language. 1993; 32:115–140.

Giegerich, H. English Phonology. Cambridge: Cambridge University Press; 1992.

Gierut, Judith A. Syllable onsets: Clusters and adjuncts in acquisition. Journal of Speech, Language and Hearing Research. 1999; 42:708–726.

Goldsmith, John A. Ph.D. dissertation. MIT; 1976. Autosegmental phonology.

Gouskova, Maria. Falling sonority onsets, loanwords, and syllable contact. CLS. 2001; 37:175–185.

Gouskova, Maria. Relational hierarchies in Optimality Theory: The case of syllable of syllable contact. Phonology. 2004; 21:205–250.

Greenberg, Joseph H. Some generalizations concerning initial and final consonant clusters. In: Greenberg, JH.; Ferguson, CA.; Moravcsik, EA., editors. Universals of Human Language. Vol. Vol. 2. Stanford, CA: Stanford University Press; 1978. p. 243-279.

Hallé, Pierre; Segui, Juan; Frauenfelder, Ulricht; Meunier, Christine. The processing of illegal consonant clusters: A case of perceptual assimilation? Journal of Experimental Psychology: Human Perception and Performance. 1998; 24:592–608. [PubMed: 9554098]

Hayes, Bruce. New methods for studying UG in phonology; A paper presented at the 38th Meeting of the North East Linguistic Society; 2007.

Hayes, Bruce; Wilson, Colin. A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry. 2008; 39

Hooper, Joan B. An introduction to natural generative phonology. New York: Academic Press; 1976.

Kawasaki-Fukumori, Haruko. An acoustical basis for universal phonotactic constraints. Language and Speech. 1992; 35:73–86. [PubMed: 1287393]

Kiparsky, Paul. Metrical structure assignment is cyclic. LI. 1979; 10:421–442.

Lacharite, Darlene; Paradis, Carole. Addressing and disconfirming some predictions of phonetic approximation for loanword adaptation. Langues et linguistique. 2002; 28:71–91.

Leben, William. Ph.D. dissertation. MIT; 1973. Suprasegmental phonology.

Levitt, Andrea; Healy, Alice F.; Fendrich, David W. Syllable-internal structure and the sonority hierarchy: Differential evidence from lexical decision, naming, and reading. Journal of Psycholinguistic Research. 1991; 20:337–363. [PubMed: 1920168]

Loftus GH, Masson ME. Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review. 1994; 1:476–490.

Massaro, Dominic W.; Cohen, Michael M. Phonological constraints in speech perception. Perception & Psychophysics. 1983; 34:338–348. [PubMed: 6657435]

McCarthy, J. Formal problems in Semitic phonology and mophology. Doctoral dissertation MIT. New York: Garland Press; 1979. 1985

Morelli, Frida. The Phonotactics and Phonology of Obstruent Clusters in Optimality Theory. Ph.D. dissertation. University of Maryland; 1999. PhD dissertation

Moreton, Elliott. Structural constraints in the perception of English stop-sonorant clusters. Cognition. 2002; 84:55–71. [PubMed: 12062147]

Moreton, Elliott. Competition in perception grammar; Papre presented at the workshop on "Experimental Approaches to Optimality Theory (ExpOT)"; 2007.

Ohala, John J. Alternatives to the Sonority Hierarchy for Explaining Segmental Sequential Constraints. Papers from the Regional Meetings, CLS. 1990; 2:319–338.

Orden, GC Van; Pennington, BF.; Stone, GO. Word identification in reading and the promise of subsymbolic psycholinguistics. Psychological Review. 1990; 97:488–522. [PubMed: 2247539]

Oudeyer, Pierre-Yves. The self-organization of speech sounds. Journal of Theoretical Biology. 2005; 233:435–449. [PubMed: 15652151]

Parker, Stephen. Quantifying the Sonority Hierarchy. PhD dissertation. University of Massachusetts; 2002.

Pater, Joe. Bridging the gap between receptive and productive development with minimally violable constraints. In: Kager, R.; Pater, J.; Zonneveld, W., editors. Constraints in Phonological Acquisition. New York: Cambridge University Press; 2004. p. 219-244.

Peperkamp S, Vendelin I, Nakamura K. On the perceptual origin of loanword adaptations: experimental evidence from Japanese. Phonology. 2008:129–164.

Peperkamp, Sharon. Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. Cognition. 2007; 104:631–637. [PubMed: 17266948]

Peperkamp, Sharon; Dupoux, Emmanuel. Reinterpreting loanword adaptations: The role of perception. In: Beachley, B.; Brown, A.; Conlin, F., editors. Proceedings of the 27th Annual Boston University Conference on Language Development; Sommerville, MA: Cascadilla Press; 2003. p. 650-661.

Pertz, Doris L.; Bever, Thomas G. Sensitivity to phonological universals in children and adolescents. Lg. 1975; 51:149–162.

Pinker, Steven; Birdsong, David. Speakers' sensitivity to rules of frozen word order. Journal of Verbal Learning and Verbal Behavior. 1979; 18:497–508.

Pitt, Mark A. Phonological processes and the perception of phonotactically illegal consonant clusters. Perception & Psychophysics. 1998; 60:941–951. [PubMed: 9718954]

Prince, A.; Smolensky, P. Optimality theory: Constraint interaction in generative grammar. Malden, MA: Blackwell Pub; 1993/2004.

Romani, Cristina; Calabrese, Andrea. Syllabic constraints on the phonological errors of an aphasic patient. Brain and Language. 1998; 64:83–121. [PubMed: 9675045]

Saussure, Ferdinand de. Course in general linguistics. New York: Philosophical library; 1915/1959.

Selkirk, Elisabeth O. On the major class features and syllable theory. In: Aronoff, Mark; Oerhle, Richrad T., editors. Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students. Cambridge, MA: MIT Press; 1984. p. 107-136.

Silverman, Daniel. Multiple Scansions in Loanword Phonology: Evidence from Cantonese. Phonology. 1992:289–328.

Smolensky, Paul. Optimality in phonology II: Markedness, feature domains, and local constraint conjunction. In: Smolensky, P.; Legendre, G., editors. The harmonic mind: From neural computation to optimality-theoretic grammar Vol.2. Linguistic and Philosophical Implications. Cambridge, MA: MIT Press; 2006. p. 27-160.

Smolensky, Paul. On the comprehension/production dilemma in child language. Linguistic Inquiry. 1996; 27:720–731.

Stemberger, Joseph; Treiman, Rebecca. The internal structure of word-initial consonant clusters. Journal of Memory and Language. 1986; 25:163–180.

Stenneken, Prisca; Bastiaanse, Roelien; Huber, Walter; Jacobs, Arthur M. Syllable structure and sonority in language inventory and aphasic neologisms. Brain and language. 2005; 95:280–292. [PubMed: 16246736]

Steriade, Donca. Greek prosodies and the nature of syllabification; PhD dissertation. MIT; 1982. PhD dissertation

Steriade, Donca. Reduplication and syllable transfer in Sanskrit and elsewhere. Phonology. 1988; 5:37–155.

Surprenant, Aimee M.; Goldstein, Louis. The perception of speech gestures. JASA. 1998; 104:518–529.

Tesar, Bruce. An iterative strategy for learning metrical stress in Optimality Theory. In: Hughes, E.; Hughes, M.; Greenhill, A., editors. Proceedings of the 21st Annual Boston University Conference on Language Development; Somerville, MA: Cascadilla; 1997. p. 615-626.

Tesar, Bruce. An iterative strategy for language learning. Lingua. 1998; 104

Tesar, Bruce. Robust interpretive parsing in metrical stress theory. In: Shahin, K.; Blake, S.; Kim, E-S., editors. Proceedings WCCFL 17; Stanford, CA: CSLI; 1999. p. 625-639.

Treiman, Rebecca. On the status of final consonant clusters in English syllables. Journal of Verbal Learning and Verbal Behavior. 1984; 23:343–356.

Treiman, Rebecca; Bowey, Judith A.; Bourassa, Derrick. Segmentation of spoken words into syllables by English-speaking children as compared to adults. Journal of Experimental Child Psychology. 2002; 83:213–238. [PubMed: 12457860]

Treiman, Rebecca; Danis, Catalina. Syllabification of intervocalic consonants. Journal of Memory and Language. 1988; 27:87–104.

Vennemann, Theo. On the theory of syllable phonology. Linguistische Berichte. 1972; 18:1–18.

Wright, Richard. A review of perceptual cues and robustness. In: Steriade, D.; Kirchner, R.; Hayes, B., editors. Phonetically based phonology. Cambridge: Cambridge University Press; 2004. p. 34-57.

Yip, Moira. The symbiosis between perception and grammar in loanword phonology. Lingua. 2006; 116:950–975.

Zec, Draga. The syllable. In: de Lacy, Paul, editor. The Cambridge Handbook of Phonology. Cambridge: Cambridge University Press; 2007. p. 161-194.

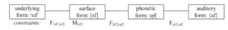

Fig. 1.
The proposed phonological processing architecture.

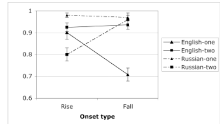

Figure 2.
Mean response accuracy of English and Russian speakers in the syllable count task. Error bars reflect confidence intervals, constructed for the difference between the means.
Note. “One” and “two” represent monosyllabic and disyllabic items, respectively.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

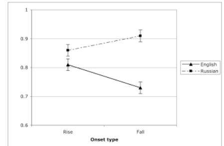

Figure 3.
Mean response accuracy of English and Russian speakers to nonidentical items. Error bars reflect 95% confidence intervals, constructed for the difference between the means.

Table 1

Mean Response Time and Accuracy in Experiment 1.

	Response accuracy (% Correct)		Response time (in ms)	
	Sonority rise	**Sonority fall**	**Sonority rise**	**Sonority fall**
Monosyllabic items	90.1	70.8	1273	1300
Disyllabic items	92.4	93.6	1328	1320

Table 2

Mean response Time and Accuracy to the nonidentity items in Experiment 2.

Onset type	Sonority rise (e.g., *mlif*)	Sonority fall (e.g., *mdif*)
Response accuracy (% Correct)	80.7	72.6
Response time (in ms)	1119.5	1121.2

Table 3

Mean response time of Russian and English speakers in the syllable count task.

	Monosyllabic items		Disyllabic items	
	Sonority Rise (e.g., *mlif*)	**Sonority Fall (e.g., *mdif*)**	**Sonority Rise (e.g., *mlif*)**	**Sonority Fall (e.g., *mdif*)**
English speakers	1273	1300	1327	1320
Russian speakers	1402	1432	1463	1430

Table 4

Mean response time of Russian and English speakers to nonidentical items.

	Onset type	
Language	**Sonority Rise (e.g., *mlif*)**	**Sonority Fall (e.g., *mdif*)**
English speakers	1119	1125
Russian speakers	1300	1305

Table 5

Mean response time and accuracy to nonidentity trials in Experiment 6.

	Onset type	
	Sonority rise (e.g., *mlif*)	**Sonority Fall (e.g., *mdif*)**
Response accuracy (% Correct)	90.6	85.1
Response time (in ms)	774	778

Table 6

The unique effects of sonority and statistical properties entered as the last predictors in linear stepwise regression analyses of Experiments 1–3 & 6

Experiment	Last Predictor	R^2 Change	Df	F	P value
1	Sonority	0.110	1,18	3.57	0.06
Syllable count	Statistical properties				
	(word & neighborhood)	0.093	4,18	<1	
(auditory)	Sonority	0.212	1,18	13.67	0.003
	Statistical properties				
	(onset & neighborhood)	0.274	4,18	4.42	0.02
2	Sonority	0.063	1,18	1.93	.19
Identity task	Statistical properties				
(auditory)	(word & neighborhood)	.173	4, 18	1.33	0.30
	Sonority	0.111	1,18	4.22	0.06
	Statistical properties				
	(onset & neighborhood)	0.284	4, 18	2.69	0.07
6	Sonority	0.182	1,18	5.94	0.03
Identity task	Statistical properties				
(visual)	(word & neighborhood)	0.204	4, 18	1.67	0.21
	Sonority	0.085	1,19	2.56	0.13
	Statistical properties				
	(onset & neighborhood)	0.119	3, 19	<1	
3	Sonority	0.149	1, 18	4.75	.05
Transcription	Statistical properties				
(off-line)	(word & neighborhood)	0.381	4, 18	3.04	.05
	Sonority	0.001	1, 18	<1	
	Statistical properties				
	(onset & neighborhood)	0.439	4, 18	3.89	.02

Note: Word and neighborhood properties include the number of neighbors, neighbor frequency and either phoneme and biphone probability (for auditory words) or bigram count and frequency (for printed words). Onset and neighborhood properties include onset probability, the number of neighbors, and neighbor frequency; the analyses on auditory words also include the initial phoneme (*m* or *n*).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 7

The effect of sonority profile on response accuracy (% correct) to *m*- vs. *n*-initial onsets.

	Onset type			
	ml	***md***	***nw***	***nb***
Experiment 1	92.70	61.00	87.30	80.80
Experiment 2	81.70	67.50	79.50	77.30
Experiment 3	64.46	55.21	39.58	27.08
Experiment 6	91.50	82.50	89.80	87.50

6. Appendix

Sonority rise	Sonority fall
mlɪf	mdɪf
mlɛf	mdɛf
mlæk	mdæk
mlɛb	mdɛb
mlʌp	mdʌp
mlɛk	mdɛk
nwɑt	nbɑt
nwɪk	nbɪk
nwɛf	nbɛf
nwɑg	nbɑg
nwʌf	nbʌf
nwɑd	nbɑd

NIH-PA Author Manuscript
NIH-PA Author Manuscript