

# Phonological Universals Constrain the Processing of Nonspeech Stimuli

Iris Berent  
Northeastern University

Evan Balaban  
McGill University

Tracy Lennertz  
Northeastern University

Vered Vaknin-Nusbaum  
University of Haifa and Western Galilee College

Domain-specific systems are hypothetically specialized with respect to the outputs they compute and the inputs they allow (Fodor, 1983). Here, we examine whether these 2 conditions for specialization are dissociable. An initial experiment suggests that English speakers could extend a putatively universal phonological restriction to inputs identified as nonspeech. A subsequent comparison of English and Russian participants indicates that the processing of nonspeech inputs is modulated by linguistic experience. Striking, qualitative differences between English and Russian participants suggest that they rely on linguistic principles, both universal and language-particular, rather than generic auditory processing strategies. Thus, the computation of idiosyncratic linguistic outputs is apparently not restricted to speech inputs. This conclusion presents various challenges to both domain-specific and domain-general accounts of cognition.

*Keywords:* phonological universals, phonology, nonspeech, optimality theory, speech perception

The nature of the human capacity for language is one of the most contentious issues in cognitive science. The debate specifically concerns the specialization of the language system—whether people are innately equipped with mechanisms dedicated to the computation of linguistic structure (e.g., Chomsky, 1980; Pinker, 1994) or whether language processing relies only on domain-general systems (Elman et al., 1996; McClelland, 2009; Rumelhart & McClelland, 1986).

Of the various behavioral hallmarks of specialization, the test from design is arguably the strongest. If language were the product of a specialized system, then one would expect all languages to share universal design principles (Chomsky, 1980; Fodor, 1983; Jackendoff, 2002; Pinker, 1994). Moreover, if the language system emerged by natural selection, these principles should be functionally adaptive (Pinker, 2003; Pinker & Bloom, 1994). To the extent that such universal adaptive principles exist and they are demonstrably robust with respect to the statistical properties of linguistic experience and input modality, this would provide strong suggestive evidence for constraints inherent to the language system itself (but cf. Blevins, 2004; Bybee, 2008; Evans & Levinson, 2009;

Kirby, 1999). However, such universal structural restrictions would guarantee that the language system is constrained only with respect to its outputs. It has previously been proposed that securing the specialization of the system also requires narrow constraints on what qualifies as input. A specialized language system should admit linguistic stimuli but deny access to nonlinguistic inputs. As Fodor (1983) put it, an input analyzer is domain-specific inasmuch as “only a relatively restricted class of stimulations can throw the switch that turns it on” (p. 49). Together, these two conditions—the unique structural constraints on the system’s outputs and its selectivity with respect to its inputs—offer a strong test of specialization by design.

In what follows, we do not attempt to directly determine whether the language system is, in fact, specialized. Instead, we examine whether the two conditions of specialization—the specificity of outputs and inputs—can diverge on their outcomes: Are the same principles that appear to be universal well-formedness conditions on linguistic outputs also applied to nonlinguistic inputs? Our investigation specifically concerns the phonological component of the grammar.

To this end, we scrutinize the sonority sequencing principle (e.g., Clements, 1990; Greenberg, 1978). This principle has been widely documented across languages, and previous experimental research has suggested that English speakers extend this knowledge even to structures that are unattested in their language (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007). The generality of this principle across language modalities, both spoken and signed (Corina, 1990; Sandler, 1993; Sandler & Lillo-Martin, 2006), and its robustness with respect to linguistic experience (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008) render it a particularly strong candidate for a universal restriction on linguistic outputs. Surprisingly, however, the relevant knowledge is not selective with re-

---

Iris Berent and Tracy Lennertz, Department of Psychology, Northeastern University; Evan Balaban, Department of Psychology, McGill University, Montreal, Quebec, Canada; Vered Vaknin-Nusbaum, Faculty of Education, University of Haifa, Haifa, Israel, and Department of Education, Western Galilee College, Akko, Israel.

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01DC003277. We wish to thank Alex Fine for statistical consultation.

Correspondence concerning this article should be addressed to Iris Berent, Department of Psychology, Northeastern University, 125 Nightingale Hall, 360 Huntington Avenue, Boston, MA 02115-5000. E-mail: i.berent@neu.edu

spect to its inputs. In the present series of experiments, we demonstrate that English speakers extend the relevant generalization to a wide range of auditory inputs, irrespective of whether these stimuli are identified as speechlike. A comparison of English speakers to speakers of Russian indicates that the processing of inputs identified as nonspeech is modulated by linguistic experience. The striking, qualitative differences between the two groups suggest that the processing of nonspeech inputs relies on linguistic knowledge, including principles that are both universal and language-particular. Although these results do not unequivocally establish whether the relevant restrictions are domain-specific, they open up the possibility that the computation of idiosyncratic linguistic outputs might not be restricted to speech inputs. This apparent dissociation between output specificity and input specificity presents multiple challenges to both domain-specific and domain-general accounts.

We begin the discussion by reviewing our case study—the restrictions on the sonority sequencing of onset consonants. We next examine whether these broad restrictions on phonological outputs also apply only to speech inputs.

### Sonority Restrictions as a Putative Output Universal

All spoken languages constrain the co-occurrence of phonemes in the syllable. For example, syllables like *bla* are universally preferred to syllables like *lba*. Not only are *bla*-type syllables frequent across languages, but any language that allows the *lba* type is likely to allow the *bla* type as well (Greenberg, 1978). Linguistic research attributes such facts to universal grammatical restrictions concerning the sonority of segments (e.g., Clements, 1990; Kiparsky, 1979; Parker, 2002; Selkirk, 1984; Steriade, 1982; Zec, 2007; see also Hooper, 1976; Saussure, 1915/1959; Vennemann, 1972). Sonority is a scalar phonological property that correlates with the intensity of consonants: least sonorous (i.e., softest) are stops (e.g., *b*, *d*) with a sonority (*s*) level of 1 ( $s = 1$ ), followed by nasals (e.g., *m*, *n*;  $s = 2$ ) and liquids (e.g., *l*, *r*;  $s = 3$ ). The contrast between *bla* and *lba* specifically concerns the sonority of their onset—the sequence of consonants that occur prior to the vowel. Syllables such as *bla* manifest a rise in sonority from the obstruent *b* ( $s = 1$ ) to the liquid *l* ( $s = 3$ ), an increase of two steps ( $\Delta s = 2$ ), whereas in *lba*, the onset falls in sonority ( $\Delta s = -2$ ). The frequency of *bla*-type syllables might thus reflect a broader preference for onsets to rise in sonority—the larger the rise, the better formed the syllable.

Typological research shows that onsets with small sonority clines are systematically underrepresented across languages, and any language that tolerates such small clines is reliably more likely to exhibit larger clines (Greenberg, 1978; analyzed in Berent et al., 2007). Specifically, languages that allow sonority falls (e.g., *lba*,  $\Delta s = -2$ ) tend to allow sonority plateaus (e.g., *bda*,  $\Delta s = 0$ ); languages that allow sonority plateaus tend to allow small sonority rises (e.g., *bna*,  $\Delta s = 1$ ); and languages that allow small sonority rises allow larger rises (e.g., *bla*,  $\Delta s = 2$ ). These results suggest that onsets with large sonority clines are preferred to those with smaller clines, which, in turn, are preferred to sonority plateaus. Least preferred are onsets that fall in sonority.

A large body of experimental research indicates that onsets with large sonority clines are indeed preferable (for reviews, see Berent et al., 2009, 2007; but cf. Davidson, 2006a, 2006b; Peperkamp,

2007; Redford, 2008), but because such onsets also tend to be more frequent in language use, the precise source of such preferences—sonority or frequency—is difficult to ascertain. A recent series of experiments, however, demonstrated that speakers favor onsets with large sonority clines to those with smaller clines even when both types of onsets are unattested in their language (Berent, 2008; Berent et al., 2008, 2009, 2007). Such preferences were inferred from the susceptibility of unattested onsets to undergo perceptual repair.

It is well known that ill-formed onsets tend to be misidentified (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; Hallé, Segui, Frauenfelder, & Meunier, 1998; Massaro & Cohen, 1983). For example, English speakers misidentify the unattested onset *tla* as *tela* (Pitt, 1998). Interestingly, however, not all onsets are equally likely to be misidentified. Instead, the rate of misidentification depends on the sonority structure of the onset. Specifically, onsets with sonority falls are more likely to be misidentified (e.g., *lba* → *leba*) compared with onsets having sonority plateaus, which, in turn, are more susceptible to misidentification than are onsets with small sonority rises (Berent et al., 2008, 2007). Several observations suggest that the systematic misidentification of onsets with small sonority clines is not simply due to an inability to encode their acoustic properties, as speakers are demonstrably able to discriminate ill-formed onsets from their disyllabic counterparts (e.g., *lba* vs. *leba*) when they attend to their surface phonetic form (Berent, Lennertz, & Balaban, 2010; Berent et al., 2007). Moreover, the processing difficulty of ill-formed onsets obtains even when they are presented in print (Berent & Lennertz, 2010; Berent et al., 2009). Speakers' behavior is also inexplicable by the statistical properties of the English lexicon (Berent et al., 2009, 2007). Indeed, similar preferences were observed even among speakers of Korean—a language that lacks any onset clusters altogether (Berent et al., 2008).

A systematic dispreference for onsets with small sonority clines could reflect a universal grammatical restriction that is active in the brains of all speakers. This restriction disfavors onsets with small sonority clines (Smolensky, 2006). Accordingly, onsets such as *lba* are worse formed than those such as *bda*, which, in turn, are worse formed than those such as *bna*. The misidentification of onsets with small sonority clines is directly due to their grammatical ill-formedness (Berent et al., 2007). In this view, ill-formedness prevents the faithful encoding of onsets such as *lba* by the grammar and leads to their recoding as better formed structures (as *leba*)—the smaller the cline, the more likely the recoding (for a formal account, see Berent et al., 2009). Like all grammatical constraints, however, the constraint on sonority is violable (de Lacy, 2006; McCarthy, 2005; Prince & Smolensky, 1997, 1993/2004; Smolensky & Legendre, 2006), and consequently, highly ill-formed onsets of falling sonority are nonetheless attested in some languages. But if the constraint on sonority is universally encoded in all grammars, then any language that allows *lba*-type onsets will also allow for better formed onsets. Similarly, given the choice, grammatical processes will be less likely to preserve *lba*-type onsets even in languages in which such onsets are attested. The typological and linguistic facts are consistent with this prediction.

Further support for the view of sonority restrictions as a universal grammatical constraint is offered by their phonetic ground-

ing and their generality across language modalities. Several authors have argued that phonological restrictions have a strong phonetic basis (e.g., Hayes, Kirchner, & Steriade, 2004, and references therein). This claim does not eliminate the demonstrable need for encoding such restrictions in the phonological grammar. Rather, it offers an explanation as to why such constraints could present an adaptive advantage that could have favored their grammaticalization during language evolution. In this respect, it is interesting that sonority restrictions have a strong acoustic (Ohala, 1990; Wright, 2004) and articulatory (Mattingly, 1981) basis. Nonetheless, sonority restrictions are not restricted to the spoken modality, as similar constraints are observed in the representation of sign language (Corina, 1990; Sandler, 1993; Sandler & Lillo-Martin, 2006). As in the spoken language, the sonority of signs correlates with their perceptual salience, and syllables that rise in sonority are preferred. The linguistic documentation of adaptive sonority restrictions in a wide variety of languages, both spoken and signed, and the experimental support of these constraints among speakers who lack the relevant structures in their language suggest that sonority restrictions are a strong candidate for a universal domain-specific output condition on phonological representations.

### Does Phonological Knowledge Apply to Nonspeech Inputs?

With a plausible candidate for a universal output restriction at hand, we can now proceed to examine whether this principle is selective with respect to its inputs—whether sonority restrictions apply narrowly to linguistic inputs or also extend to stimuli that are not linguistic in nature.

Previous research has documented various differences between the phonetic processing of natural speech and various analogues that are not identified as speech (hereafter, *nonspeech*). For example, sine-wave speech analogues that are initially perceived as nonspeech by listeners are perceptually less cohesive than synthetic speech (Remez, Pardo, Piorkowski, & Rubin, 2001) and are less likely to elicit categorical perception (Mattingly, Liberman, Syrdal, & Halwes, 1971), promote audiovisual integration (Kuhl & Meltzoff, 1982; Tuomainen, Andersen, Tiippana, & Sams, 2005), or engage language regions in the brain (e.g., Schofield et al., 2009; Scott, Blank, Rosen, & Wise, 2000; Vouloumanos, Kiehl, Werker, & Liddle, 2001). Nonetheless, after minimal experience, such nonspeech stimuli can give rise to phonetic processing (e.g., trading relations; Best, Morrongoello, & Robson, 1981), allowing for the identification of linguistic messages (Remez et al., 2001; Remez, Rubin, Berns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981) and engaging phonetic brain networks (Liebenthal, Binder, Piorkowski, & Remez, 2003; Meyer et al., 2005).

The finding that the phonetic system can be engaged by stimuli that are not invariably perceived as speech suggests that the language system is not narrowly restricted with respect to its input (Remez et al., 1994). The scope of such restrictions, however, is not entirely clear from the present literature. Several authors have argued that the phonetic system is engaged only for stimuli that are consciously identified as speech (Best et al., 1981; Liebenthal et al., 2003; Meyer et al., 2005; Remez et al., 1981; Tuomainen et al., 2005). However, Azadpour and Balaban (2008) observed that

discrimination between spectrally rotated vowels—stimuli identified by all participants as nonspeech—is reliably affected by their phonetic distance even after statistically controlling for their acoustic similarity. These findings suggest that the phonetic system can be engaged even by stimuli that are not consciously identified as speech (Azadpour & Balaban, 2008). Another open question concerns the processing depth of nonspeech stimuli. Although existing research has shown that nonspeech stimuli can undergo various aspects of phonetic processing, no previous study has examined whether such stimuli can trigger knowledge of broad phonological principles.

The present research investigates these questions. We first examine whether English speakers extend the restriction on sonority to auditory stimuli they have identified as either speech or nonspeech. Because these stimuli include onsets that are unattested in English, the sensitivity of English speakers to the sonority of such onsets potentially reflects the generalization of universal phonological principles to nonspeech. We next investigate whether nonspeech stimuli also activate language-particular knowledge. To this end, we compare English speakers with speakers of Russian—a language that allows such onsets—given both speech and nonspeech stimuli. Finally, in the General Discussion section, we gauge what type of knowledge is consulted by participants.

### Part 1: Do English Speakers Extend Sonority Knowledge to Nonspeech?

Experiments 1–2 investigated whether English speakers extend sonority restrictions to synthetic stimuli—either speechlike stimuli or items that do not sound like speech. Our procedure follows previous research with natural speech stimuli (Berent et al., 2010), which examined the identification of speech continua, ranging from  $C_1C_2VC_3$  monosyllables ( $C$  = consonant,  $V$  = vowel) to their disyllabic  $C_1\sigma C_2VC_3$  counterparts. In one such continuum, the CCVC monosyllable had an onset of rising sonority (e.g., *mlif*), whereas in the second continuum, the monosyllable had an onset that fell in sonority (e.g., *mdif*).

Stimuli along these two continua were generated by a procedure of incremental splicing, along the lines described in Dupoux et al. (1999). First, we had an English talker naturally produce the two disyllabic end points (e.g., *məlIf*, *mədIf*) and selected tokens that were matched for the duration of the pretonic vowel (a schwa). For each such disyllable, we next gradually excised the pretonic vowel in five steady increments, resulting in a continuum of six steps, ranging from the fully monosyllabic form (Step 1) to the disyllabic end point (Step 6).

Our experiments compared the responses of English participants to the *ml* and *md* continua. In each trial, participants heard one item, randomly selected from the two continua. We asked them to determine whether the stimulus included “one beat” or “two beats” and illustrated the task using English words (e.g., *sport* vs. *support*, as examples of one vs. two beats, respectively). Results showed, that, as the duration of the pretonic vowel increased, participants were more likely to identify the stimulus as having two beats, and given a fully disyllabic form in Step 6, the rate of two-beat responses did not differ for the rise and fall continua. Crucially, for each of the other five steps, participants were reliably more likely to identify the input as having two beats for the fall compared with the rise continua. This result indicates that onsets of falling sonor-

ity are represented as disyllabic, a finding that might be due to their grammatical ill-formedness.

In the present research, we examined whether this same pattern of results extends to nonspeech inputs. Nonspeech materials were generated by resynthesizing the first formant of the natural speech continua used in our previous research in a manner that renders them distinctly nonspeechlike (see Figure 1A)<sup>1</sup>. To partly control for any editing artifacts, we compared these nonspeech items with a filtered version of the original stimuli that is typically identified as speech (see Figure 1B). Experiment 1 seeks to establish that sonority restrictions generalize to speech controls. Experiment 2 investigates whether the same constraints apply to nonspeech items.

### Experiment 1: The Sensitivity of English Speakers to the Structure of Speech-Control Stimuli

#### Method.

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

**Materials.** The materials were generated by resynthesizing the continua of natural-speech stimuli used in Berent et al. (2010). Each such continuum ranged from a monosyllable to a disyllable. Monosyllables included two types of nasal-initial onsets, manifesting either a rise or fall in sonority (e.g., /mlɪf/, /mdɪf/). These items were arranged in pairs, matched for their rhyme (i.e., the main vowel and the following consonant). There were three such pairs in the experiment (/mlɪf/-/mdɪf/, /mlɛf/-/mdɛf/, and /mlɛb/-/mdɛb/). Each item was generated by a procedure of incremental splicing, as described in Dupoux et al. (1999). We first had a native English speaker (naïve to this research project) naturally produce the disyllabic counterparts (e.g., /məlɪf/ and /mədɪf/) and then selected pairs that were matched for total length, intensity, and the duration of the pretonic schwa (68 ms for both /məlɪf/- and /mədɪf/-type items). We next continuously extracted the pretonic vowel at the zero crossings in five steady increments, moving from its center outward. This procedure yielded a continuum of six steps, ranging from the original disyllabic form (e.g., /məlɪf/) to an onset cluster in which the pretonic vowel was fully removed (e.g., /mlɪf/). The number of pitch periods in Stimuli 1–5 was 0, 2, 4, 6, and 8, respectively; Stimulus 6 (the original disyllable) ranged from 12 to 15 pitch periods.

These natural-speech continua were next resynthesized, using a digital filter with a slope of -85 dB per octave above the stated cutoff frequency (1216 Hz for /məlɪf/- and /mədɪf/-type items, 1270 Hz for /mɛləf/-type items, 1110 Hz for /mɛləb/-type items, 1347 Hz for /mɛdɛf/-type items, and 1250 Hz for /mɛdəb/-type items), designed to reduce the information available at higher frequencies. This does not completely remove the higher frequency information, because the human ear is extremely good at detecting speech information in the “shoulders” of the filter (for speech stimuli, see Warren, Bashford, & Lenz, 2004, 2005). This manipulation was done as a control manipulation to acoustically alter the stimuli in a manner similar to that for the nonspeech stimuli described later in Experiment 2, yet it preserves enough speech information for these items to be identified as (degraded) speech.

The six-step continuum for each of the three pairs was presented in all six durations, resulting in a block of 36 trials. Each such

block was repeated four times, yielding a total of 144 trials. The order of trials within each block was randomized.

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

After the experiment, participants were presented with two questions, designed to determine whether the resynthesized stimuli were, in fact, identified as speech. The two questions were (a) “Can you describe the sounds you have heard in this experiment?” and (b) “Do they sound like speech sounds?” Participants responded in writing.

**Results and discussion.** Most (11 out of 12) participants described the stimuli using terms related to speech or language (e.g., baby talk, blurred words, small child learning to speak), and when explicitly asked whether the stimuli sounded like speech, seven of the 12 participants responded “yes.”

Figure 2 plots the proportion of two-beat responses as a function of vowel duration and continuum type (sonority rise vs. fall). In this and all subsequent figures, error bars correspond to the 95% confidence intervals, constructed for the difference between the mean responses to the rise and fall continua at each of the six steps (e.g., the difference between *mlif* and *mdif* at Step 1) using the pooled mean square error across all six steps. An inspection of the means suggests that, as vowel duration increased, people were more likely to identify the input as including two beats (i.e., disyllabic). However, the rate of two-beat responses was higher for the fall continuum compared with the rise continuum.

A 2 (continuum type) × 6 (vowel duration) analysis of variance (ANOVA) yielded significant main effects of continuum type,  $F(1, 11) = 357.01, p < .0001, \eta^2 = 0.97$ ; vowel duration,  $F(5, 55) = 53.99, p < .0001, \eta^2 = 0.83$ ; and their interaction,  $F(5, 55) = 28.06, p < .0001, \eta^2 = 0.72$ .<sup>2</sup> The simple main effect of vowel duration was significant for both the rise,  $F(5, 55) = 55.19, p < .0002, \eta^2 = 0.83$ , and fall,  $F(5, 55) = 2.74, p < .03, \eta^2 = 0.20$ , continuum, indicating that, as vowel duration increased, people

<sup>1</sup> We chose to preserve the first formant because it carries important cues that correlate with the sonority of our materials. Specifically, the structure of the first formant contrasts liquids and stops, and the zero-pole pattern indicates nasality (Wright, 2004).

<sup>2</sup> Binary data, such as a binary syllable count, could lead to artifacts in the ANOVA (Jaeger, 2008). Although multilevel logit models address these concerns, they are not entirely optimal for our present design—a design that critically hinges on the interpretation of numerous high-order interactions. Because the statistical mechanisms necessary to draw such inferences are still not fully developed in logit models, we decided to adopt a hybrid approach that builds on the strengths of both perspectives: We maintain the ANOVA in order to interpret higher order interactions, but as additional assurance against artifacts, we also report the test of the relevant interactions using a multilevel logit model. The results are reported in the Appendix.

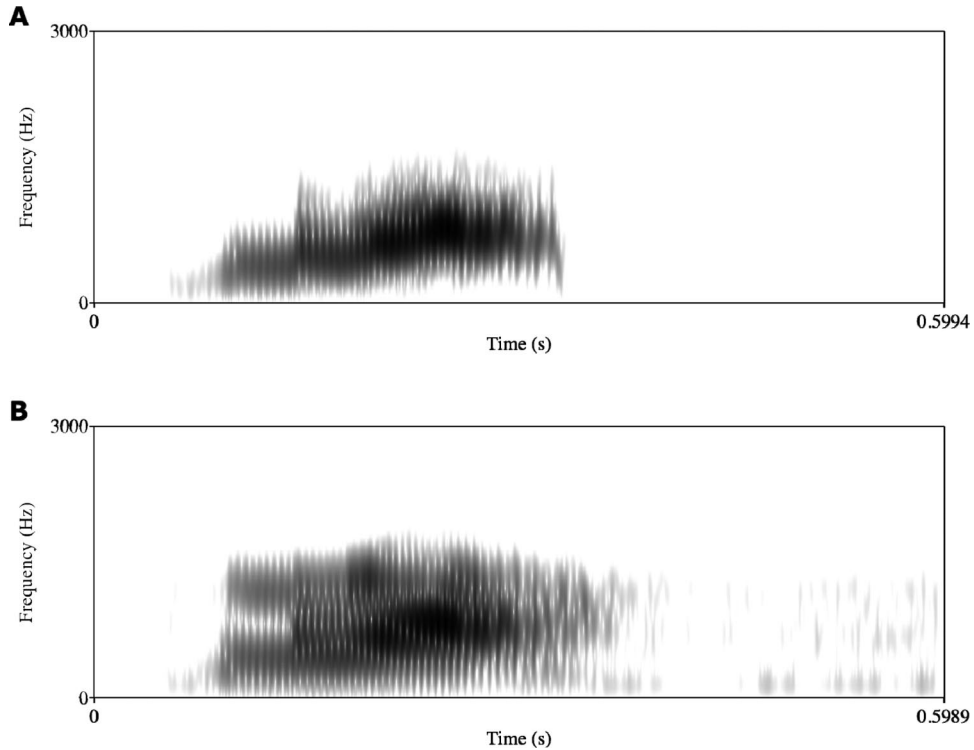


Figure 1. A spectrogram of the nonspeech (A) and speech-control (B) stimuli for *mlif* (in Step 1).

were more likely to identify the stimulus as having two beats. At each of the six vowel durations, however, two-beat responses were significantly more frequent for the fall relative to the rise continuum. The simple main effect of continuum type was significant for

Step 1,  $F(1, 11) = 149.82, p < .0002, \eta^2 = 0.93$ ; for Step 2,  $F(1, 11) = 400.77, p < .0002, \eta^2 = 0.97$ ; for Step 3,  $F(1, 11) = 268.70, p < .0002, \eta^2 = 0.96$ ; for Step 4,  $F(1, 11) = 86.56, p < .0002, \eta^2 = 0.89$ ; for Step 5,  $F(1, 11) = 65.08, p < .0002, \eta^2 = 0.86$ ; and for Step 6,  $F(1, 11) = 12.41, p < .005, \eta^2 = 0.53$ .

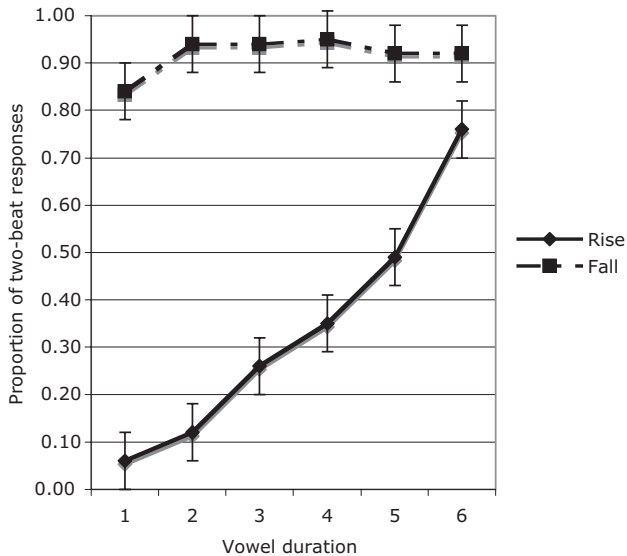


Figure 2. The proportion of two-beat responses to speech-control stimuli by English speakers in Experiment 1. Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

These results replicate our previous findings with natural speech (Berent et al., 2010) and extend them to synthetic stimuli. The only remarkable difference is that, unlike natural speech, the synthetic fall continuum elicited a higher rate of two-beat responses even at the disyllabic end point. We suspect that the unfamiliarity with synthetic input promoted greater attention to acoustic correlates of disyllabicity and that the slight acoustic discontinuities associated with splicing promoted a disyllabic percept. Because the end points were unspliced, their greater continuity could have attenuated their identification as disyllabic, putting them at a disadvantage relative to spliced inputs. Although synthesis appeared to have promoted attention to phonetic detail, results from synthetic and natural speech generally converged to suggest that sonority falls tend to be misidentified as disyllabic. Grammatical accounts might attribute such misidentifications to the recoding of ill-formed onsets by the grammar (Berent et al., 2008, 2007), whereas domain-general accounts might assert that sonority falls are harder to encode (Blevins, 2004; Bybee, 2008; Evans & Levinson, 2009). Here, we will not attempt to adjudicate between these explanations. Instead, our goal is to determine whether the same findings extend to nonspeech stimuli. Experiment 2 addresses this question. The experiment is identical to Experiment 1, with the only difference that the materials are now synthesized in a manner that renders them nonspeechlike.

### Experiment 2: The Sensitivity of English Speakers to the Structure of Nonspeech Stimuli

#### Method.

**Participants.** Twelve native English speakers, students at Florida Atlantic University, took part in the experiment in partial fulfillment of a course requirement. None of these participants took part in Experiment 1.

**Materials.** The materials were generated by resynthesizing the same set of natural recordings used in Experiment 1. The only difference was that the resynthesis method was now designed to generate materials that were typically perceived as nonspeech. A digital spectrogram (sampling rate: 22.05 kHz, DFT [discrete Fourier transform] length: 256 pts, time increment between successive DFT spectra: 0.5 ms, Hanning window) was made of each natural-speech stimulus using the SIGNAL digital programming language (Engineering Design, Berkeley, CA). Spectral contour detection was used to extract a spectral contour of the first formant by specifying an initial starting frequency that matched the center frequency of the first formant (measured by moving a screen cursor to the center of the first formant on the spectrogram of each sound). A peak-picking algorithm then found the peak in the spectral envelope of the first DFT spectrum closest to this frequency with energy that was no more than 45 dB below the global signal maximum. This value was then used to find the closest peak to it in the spectral envelope of the successive spectrum that was no more than 45 dB below the global signal maximum, and this process was repeated with successive spectra until the end of the sound was reached. Time intervals with no spectral peak greater than 45 dB below the signal maximum were coded as silent, so that silent gaps in the first formant were represented as silent gaps in the spectral contour. The amplitude values corresponding to each of the frequency values of the spectral contour were next recovered from the sound spectrogram. These amplitude and frequency contours, like the formants they were derived from, exhibit modulations in their values over time, so that a sound resynthesized from them will have a spectrogram that will not look like a pure-tone sound. These contours were turned back into a sound by digitally resampling them to the same sampling frequency as the original sound using cubic-spline interpolation and then using a digital implementation of a voltage-controlled oscillator to turn the frequency (spectral) contour into a frequency-modulated sine wave. This output was then multiplied by the amplitude contour to produce a synthetic version of the first formant. The output amplitude of this final signal was then adjusted to approximate the level of the original stimulus.

**Procedure.** The procedure was identical to that in Experiment 1, with the only difference being that the materials were now introduced as “artificial sounds generated by a computer.” Participants were informed that the stimuli were “designed to imitate a scrubbing noise, either scrubbing once or twice. One scrub includes one beat, two scrubs include two beats.”

**Results and discussion.** Participants generated a variety of descriptions of the auditory stimuli (e.g., pitches, weird alien sounds, bubbles, raindrops, cell phone bleep). No participant spontaneously used terms invoking either speech or language. Likewise, when explicitly asked whether the stimuli sounded like speech, most (9 out of 12) participants responded “no,” two responded “maybe,” and one “yes.” But despite the identification

of these stimuli as nonspeech, the pattern of results was quite similar to the one observed in Experiment 1 (see Figure 3).

F3

A 2 (continuum type) × 6 (vowel duration) ANOVA yielded significant main effects of continuum type,  $F(1, 11) = 30.90, p < .0002, \eta^2 = 0.74$ ; vowel duration,  $F(5, 55) = 9.73, p < .0001, \eta^2 = 0.47$ ; and their interaction,  $F(5, 55) = 5.68, p < .0003, \eta^2 = 0.34$ . A test of the simple main effects of vowel duration demonstrated that, as vowel duration increased, participants were reliably more likely to classify the inputs along the rise continuum as having two beats,  $F(5, 55) = 8.91, p < .002, \eta^2 = 0.45$ , and marginally so for the fall continuum,  $F(5, 55) = 2.28, p < .06, \eta^2 = 0.17$ . Moreover, participants were reliably more likely to classify the input as having two beats for items along the fall continuum compared with the corresponding steps along the rise continuum. The simple main effect of continuum type was significant for Step 1,  $F(1, 11) = 30.52, p < .0003, \eta^2 = 0.74$ ; Step 2,  $F(1, 11) = 61.18, p < .0002, \eta^2 = 0.85$ ; Step 3,  $F(1, 11) = 28.89, p < .003, \eta^2 = 0.72$ ; Step 4,  $F(1, 11) = 22.23, p < .0007, \eta^2 = 0.67$ ; Step 5,  $F(1, 11) = 18.49, p < .002, \eta^2 = 0.63$ ; and Step 6,  $F(1, 11) = 7.09, p < .03, \eta^2 = 0.39$ .

These results could indicate that people extend their phonological knowledge to inputs that they do not identify as speech. Before further considering this possibility, we must rule out the prospect that participants learned to identify the auditory stimuli as speech throughout the four blocks of the experimental session and, consequently, that the observed pattern might be based on trials that were recognized as speech. Although this possibility is inconsistent with the fact that most participants continued to identify these stimuli as nonspeech even at the end of the experimental session, we nonetheless addressed it by inspecting responses to the first encounter with each item, during the first block of trials. The results (see Table 1) closely match the pattern obtained across blocks. Specifically, a 2 (continuum type) × 6 (vowel duration) ANOVA yielded a significant interaction,  $F(5, 55) = 3.78, p <$

T1

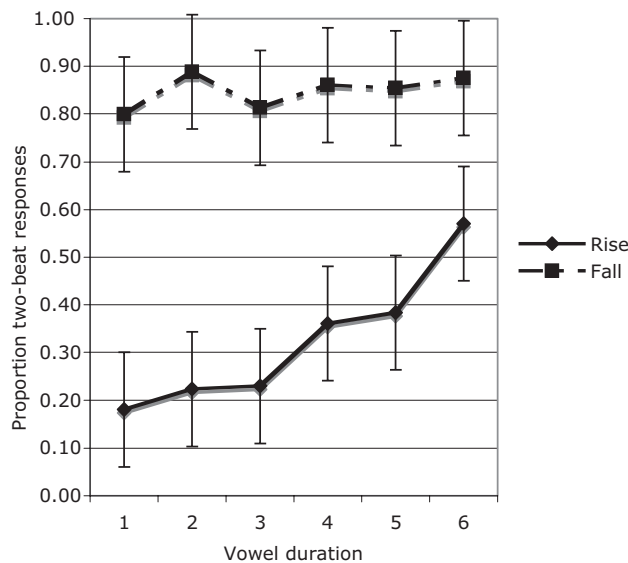


Figure 3. The proportion of two-beat responses to nonspeech stimuli by English speakers in Experiment 2. Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

Table 1  
Proportion of Two-Beat Responses by English Speakers in the First Block of Experiment 2

Vowel duration	Continuum type	
	Rise	Fall
Step 1	0.17	0.81
Step 2	0.17	0.89
Step 3	0.14	0.86
Step 4	0.22	0.89
Step 5	0.42	0.83
Step 6	0.56	0.92

.006,  $\eta^2 = 0.26$ . Vowel duration modulated responses only to the rise continuum,  $F(5, 55) = 5.78, p < .0003, \eta^2 = 0.34$ ; for the fall continuum,  $F(5, 55) < 1, \eta^2 = 0.07$ . In accord with the omnibus findings, participants were more likely to identify items along the fall continuum as having two beats in Step 1,  $F(1, 11) = 32.45, p < .0001, \eta^2 = 0.97$ ; Step 2,  $F(1, 11) = 82.14, p < .0001, \eta^2 = 0.99$ ; Step 3,  $F(1, 11) = 53.43, p < .0001, \eta^2 = 0.83$ ; Step 4,  $F(1, 11) = 65.99, p < .0002, \eta^2 = 0.86$ ; Step 5,  $F(1, 11) = 14.41, p < .005, \eta^2 = 0.57$ ; and Step 6,  $F(1, 11) = 6.78, p < .02, \eta^2 = 0.38$ . Thus, despite not having consciously identified the stimuli as speech, participants were sensitive not only to the duration of the pretonic vowel in the original speech stimuli but also to their consonant quality, and their performance with nonspeech and speech materials was quite similar.

**Part 2: Is the Perception of Nonspeech Stimuli Modulated by Linguistic Experience?**

The results of Experiments 1–2 suggest that speech and nonspeech stimuli elicit similar patterns of behavior: People are more likely to represent items along the fall continuum as having two beats. In previous research, we argued that the persistent misidentification of small sonority clines is due to their grammatical ill-formedness. The possibility that grammatical principles extend to nonspeech would suggest that the language system is nonselective with respect to its input, applying to either speech or nonspeech stimuli. However, an alternative explanation could suggest that the misidentification of items along the fall continuum is caused by the particular acoustic properties of these auditory stimuli. For example, the release closure of the stop in *md* is accompanied by a brief silent period that could promote the parsing of the acoustic inputs into two events. Because such properties are preserved in our nonspeech items, nonspeech fall continua might be more likely to elicit two-beat responses on purely acoustic grounds.

To adjudicate between these possibilities, we examine whether beat count is modulated by linguistic knowledge. Here, we do not attempt to specify the relevant knowledge—whether it is phonological or nonphonological (phonetic or even acoustic)—but only to demonstrate that the processing of synthetic speech is shaped by experience with similar linguistic stimuli. To this end, we compared the performance of English speakers with that of Russian-speaking participants. Recall that, despite the infrequency of onsets of falling sonority across languages, they are nonetheless attested in some languages. Russian is a case in point. Russian

allows nasal-initial onsets of both rising and falling sonority (e.g., *ml* vs. *mg*), albeit not the *md* onset used in our experiment. Consequently, the linguistic knowledge of Russian speakers differs from that of English speakers. If the misidentification of sonority falls is only due to the acoustic properties of such inputs, then we expect the performance of Russian speakers to resemble that of English participants. In contrast, if the misidentification of nonspeech items reflects linguistic knowledge, then, compared with English speakers, Russian speakers should be less likely to perceive the fall continuum as disyllabic. Crucially, this effect of linguistic experience should extend to the processing of both speech and nonspeech materials.

To ensure that the differences between the two groups are genuine and not confined to the processing of synthesized speech, Experiment 3 first compares the performance of Russian and English speakers with natural-speech items. Experiments 4–5 extend this investigation to synthesized speech-control and nonspeech materials, respectively.

**Experiment 3: The Sensitivity of Russian Speakers to the Structure of Natural-Speech Stimuli**

**Method.**

**Participants.** Twelve college students (native speakers of Russian) took part in this experiment. These individuals immigrated to Israel from Russia, and they were recruited at the University of Haifa in Israel. Participants were paid \$6 for taking part in the experiment.

**Materials and procedure.** The materials corresponded to the original recordings used in Berent et al. (2010), which also served as the basis for the resynthesized materials used in Experiments 1–2. The procedure was the same as in Experiment 1.

**Results and discussion.** The responses of Russian participants to natural speech from an English speaker are presented in Figure 4. An inspection of the means shows that Russian speakers F4 tended to identify items at Steps 1–5 as having one beat, regardless

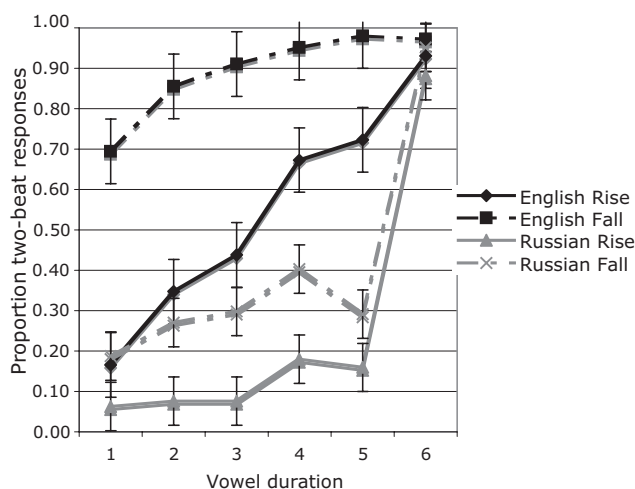


Figure 4. The proportion of two-beat responses to natural-speech stimuli by Russian speakers (in Experiment 3) and English participants (Berent et al., 2010). Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

of whether they manifested sonority rises or falls, whereas the unspliced disyllables in Step 6 were identified as having two beats.

A 2 (continuum type)  $\times$  6 (vowel duration) ANOVA yielded significant main effects of continuum type,  $F(1, 11) = 22.33, p < .0007, \eta^2 = 0.67$ , and vowel duration,  $F(5, 55) = 111.79, p < .0001, \eta^2 = 0.91$ . The interaction was marginally significant,  $F(5, 55) = 2.23, p < .07, \eta^2 = 0.17$ . Figure 4 suggests that the proportion of two-beat responses increased with vowel duration along each of the two continua. The simple main effect of vowel duration was indeed significant for both the rise,  $F(5, 55) = 173.55, p < .0002, \eta^2 = 0.94$ , and fall,  $F(5, 55) = 41.72, p < .0002, \eta^2 = 0.79$ , continua. Nonetheless, items along the fall continuum were significantly more likely to elicit two-beat responses. The simple main effect of continuum type was marginally significant in Step 1,  $F(1, 11) = 4.30, p < .07, \eta^2 = 0.28$ , and was significant in Step 2,  $F(1, 11) = 14.77, p < .003, \eta^2 = 0.57$ ; Step 3,  $F(1, 11) = 10.35, p < .009, \eta^2 = 0.48$ ; Step 4,  $F(1, 11) = 20.11, p < .001, \eta^2 = 0.65$ ; Step 5,  $F(1, 11) = 9.48, p < .02, \eta^2 = 0.46$ ; and Step 6,  $F(1, 11) = 16.18, p < .003, \eta^2 = 0.59$ . Thus, Russian speakers were more likely to give two-beat responses to items along the fall continuum, but they were overall less likely to do so than English speakers.

To establish that the responses of Russian speakers to English natural speech differ from that of English speakers, we compared the performance of the Russian participants with that of the group of English participants from Berent et al. (2010; Experiment 1). We assessed the effect of linguistic experience by means of a 2 (language)  $\times$  2 (continuum type)  $\times$  6 (vowel duration) ANOVA. Because our interest here specifically concerns the effect of linguistic experience on the perception of rise and fall continua, in this and all subsequent analyses, we discuss only the interactions involving the language factor.

The ANOVA yielded reliable interactions of Language  $\times$  Continuum Type,  $F(1, 22) = 12.47, p < .002, \eta^2 = 0.36$ , and Language  $\times$  Vowel Duration,  $F(5, 110) = 27.27, p < .002, \eta^2 = 0.55$ , as well as a significant three-way interaction,  $F(5, 110) = 5.27, p < .0003, \eta^2 = 0.19$ . To interpret this complex interaction, we proceeded to compare the responses of English and Russian participants to the rise and fall continua separately, by means of two additional 2 (language)  $\times$  6 (vowel duration) ANOVAs.

The analysis of the rise continuum yielded a significant interaction,  $F(5, 110) = 20.29, p < .0001, \eta^2 = 0.48$ , demonstrating that the effect of vowel duration was modulated by linguistic experience. Tukey HSD tests showed that English and Russian participants did not differ in their responses to the two ends of the rise continuum (both  $p > .55$ ): Both groups identified Step 1 items as including one beat, and they likewise agreed on the classification of the unspliced disyllabic stimuli at Step 6 as having two beats. But at all other steps along the continuum, English speakers were more likely to identify the input as having two beats compared with their Russian counterparts (all  $ps < .002$ ).

Similar results were obtained in the analysis of the fall continuum. Once again, linguistic experience modulated the effect of vowel duration,  $F(5, 110) = 15.90, p < .0001, \eta^2 = 0.42$ . Tukey HSD tests showed that the two groups did not differ with respect to the unspliced disyllables at Step 6 ( $p > .99$ ), but at all other steps, English speakers were reliably more likely to identify the stimulus as having two beats (all  $ps < .0002$ ). Note, however, that, unlike the responses to Step 1 of the rise continuum, linguistic

experience did modulate responses to the corresponding step on the fall spectrum: English speakers were far more likely to offer two-beat responses to sonority falls even when the pretonic vowel was altogether absent in Step 1.

In summary, Russian participants resembled their English counterparts inasmuch as both groups tended to identify items on the fall continuum as having two beats. However, Russian speakers were better able to identify the monosyllabic inputs as such, and this difference was especially noticeable for monosyllables of falling sonority. Given that linguistic experience reliably modulates responses to natural speech, we can now move to determine whether similar differences obtain for synthesized materials. Experiment 4 examines responses to speech controls; the critical comparison of these two groups with nonspeech items is presented in Experiment 5.

#### Experiment 4: The Sensitivity of Russian Speakers to the Structure of Speech-Control Stimuli

##### Method.

**Participants.** Twelve college students (native speakers of Russian) took part in this experiment. These individuals immigrated to Israel from Russia (Mean age at arrival = 11.82 years,  $SD = 6.65$ ), and they were recruited at the University of Haifa. None of these participants took part in Experiment 3. Participants were paid \$6 for taking part in the experiment.

**Materials and procedure.** The materials and procedure were the same as in Experiment 1.

**Results and discussion.** Nine of the 12 Russian participants spontaneously identified the control stimuli (generated from English speech) as speech, although most described them as speech in a foreign language. When explicitly asked whether the stimuli sounded like speech, 10 of the 12 participants responded “yes.” The remaining two responded “no,” stating that the stimuli lacked meaning or emotion, a description that does not preclude the possibility that they did, in fact, encode the input as speech. Across the two questions, 11 of the 12 participants used language or speech to describe the stimuli.

A 2 (continuum type)  $\times$  6 (vowel duration) ANOVA on the responses of the Russian participants yielded significant main effects of continuum type,  $F(1, 11) = 19.99, p < .0001, \eta^2 = 0.65$ ; vowel duration,  $F(5, 55) = 7.64, p < .0001, \eta^2 = 0.41$ ; and their interaction,  $F(5, 55) = 10.18, p < .0001, \eta^2 = 0.48$ . As with natural speech, Russian participants tended to identify stimuli on the monosyllabic end of the continua as having one beat, and they were more likely to offer one-beat responses to the rise relative to the fall continuum. Tests of the simple main effects of continuum type showed that two-beat (i.e., disyllabic) responses were significantly more frequent for the fall continuum in each of the five initial steps: Step 1,  $F(1, 11) = 16.18, p < .003, \eta^2 = 0.60$ ; Step 2,  $F(1, 11) = 15.85, p < .003, \eta^2 = 0.59$ ; Step 3,  $F(1, 11) = 32.32, p < .0002, \eta^2 = 0.73$ ; Step 4,  $F(1, 11) = 12.62, p < .005, \eta^2 = 0.53$ ; and Step 5,  $F(1, 11) = 13.10, p < .005, \eta^2 = 0.54$ . Remarkably, however, when provided with full disyllables in Step 6, Russian participants were somewhat less likely to identify the end point of the rise continuum as having two beats,  $F(1, 11) = 4.57, p < .06, \eta^2 = 0.29$ . This effect, however, should be interpreted in light of another unique finding in this experiment.



As in previous experiments, the main effects of continuum type and vowel duration were both significant. But unlike in previous experiments, the main effect of vowel duration was nonlinear: As vowel duration increased along Steps 1–5, participants were more likely to categorize the input as having two beats. Surprisingly, however, the unspliced disyllable at Step 6 was identified as monosyllabic. Tukey HSD tests confirmed that the rate of two-beat responses for Step 5 was significantly higher compared with that for Steps 1, 2, and 3 (all  $p$ s < .03), but it was also significantly higher relative to Step 6 ( $p$  < .0002). Recall that resynthesis also attenuated two-beat responses for English participants in Experiment 1, a finding we attributed to the greater reliance on acoustic cues for disyllabicity in processing the unfamiliar spliced materials. Assuming that the slight discontinuities associated with splicing promote disyllabicity, the absence of such cues in the unspliced end points would put them at a disadvantage compared with the spliced steps. For Russian speakers, this disadvantage should be exacerbated—not only are they unfamiliar with the resynthesized inputs, but the relevant linguistic structures (CəCVC) are unattested in their language. Our previous research with similar nasal-initial stimuli demonstrated that, compared with English speakers, Russian participants experience difficulty identifying CəCVC inputs as disyllabic (Berent et al., 2009). Given unfamiliar synthetic inputs with an unfamiliar disyllabic structure, Russian speakers should be more closely tuned to splicing cues than are their English-speaking counterparts, leading to a yet greater cost in the identification of the unspliced end points. These different responses of English and Russian speakers reinforce the contribution of language-particular knowledge to the interpretation of disyllables.

To further examine the effect of language-particular knowledge on responses to speech controls, we next compared the two language groups by means of a 2 (language)  $\times$  2 (continuum type)  $\times$  6 (vowel duration) ANOVA. The means are depicted in Figure 5. Replicating the results with natural speech, language experience modulated the effect of continuum type,  $F(1, 22) = 17.82, p < .005, \eta^2 = 0.45$ , and vowel duration,  $F(5, 110) = 12.98, p < .005$ ,

$\eta^2 = 0.37$ . Moreover, the three-way interaction was highly significant,  $F(5, 110) = 3.65, p < .005, \eta^2 = 0.14$ . To interpret the interaction, we next assessed the effect of linguistic experience separately for sonority rise and fall continua by means of 2 (language)  $\times$  6 (vowel duration) ANOVAs.

The analysis of the fall continuum yielded a reliable interaction of Language  $\times$  Vowel Duration,  $F(5, 110) = 6.65, p < .0001, \eta^2 = 0.23$ . Compared with their Russian counterparts, English speakers were significantly more likely to identify the input as having two beats along each of the six steps (Tukey HSD tests, all  $p$ s < .03). Language experience also modulated the responses to the rise continuum, resulting in a significant interaction,  $F(5, 110) = 13.22, p < .001, \eta^2 = 0.38$ . This interaction, however, was entirely due to the sixth and last step. A series of Tukey HSD contrasts confirmed that English and Russian speakers did not differ in their responses along the initial five steps (all  $p$ s > .38). But when provided with the full, unspliced disyllable, English speakers were more likely to interpret the input as disyllabic compared with Russian participants ( $p < .0002$ , Tukey HSD).

To summarize, Russian and English speakers generally exhibited similar responses to the rise continuum, and participants in both groups were also more likely to categorize these items as having one beat compared with items in the fall continuum. However, the two groups differed with respect to the fall spectrum: Compared with English participants, Russian speakers, whose language tolerates onsets of falling sonority, were far more likely to identify these inputs as having one beat. We now turn to examine whether linguistic experience modulates responses to synthesized stimuli that are not identified as speech.

### Experiment 5: The Sensitivity of Russian Speakers to the Structure of Nonspeech Stimuli

#### Method.

**Participants.** Twelve college students (native speakers of Russian) took part in this experiment. These individuals immigrated to Israel from Russia (Mean age at arrival = 14.04,  $SD = 6.11$ ), and they were recruited at the University of Haifa. None of these participants took part in Experiments 3–4. Participants were paid \$6 for taking part in the experiment.

**Materials and procedure.** The materials and procedure were the same as in Experiment 2.

**Results and discussion.** Six of the 12 Russian participants labeled the input using adjectives related to language or speech. When explicitly asked whether the stimuli were speechlike, six of the 12 participants responded “no.” An inspection of their responses, however, showed that they were nonetheless sensitive to the structure of these materials.

A 2 (continuum type)  $\times$  6 (vowel duration) ANOVA yielded significant main effects of continuum type,  $F(1, 11) = 16.93, p < .002, \eta^2 = 0.61$ ; vowel duration,  $F(5, 55) = 4.47, p < .002, \eta^2 = 0.29$ ; and their interaction,  $F(5, 55) = 16.81, p < .0001, \eta^2 = 0.60$ . The simple main effect of vowel duration was significant for the rise continuum,  $F(5, 55) = 5.60, p < .0004, \eta^2 = 0.34$ , reflecting a monotonic increase in the rate of two-beat responses with the increase in vowel duration. The simple main effect of continuum type was likewise significant for the fall continuum,  $F(5, 55) = 21.95, p < .0002, \eta^2 = 0.67$ .

F5

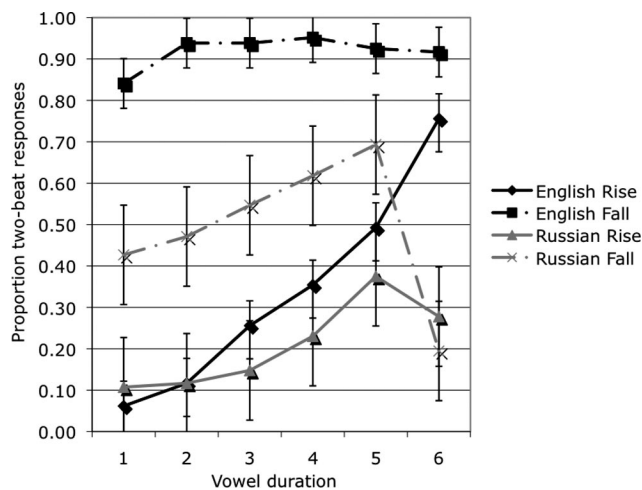


Figure 5. The proportion of two-beat responses to speech-control stimuli by Russian speakers (in Experiment 4) and English participants (in Experiment 1). Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

As with speechlike stimuli, Russian speakers were more likely to identify inputs along the fall continuum as having two beats. The simple main effect of continuum type was significant in Step 1,  $F(1, 11) = 23.71, p < .0006, \eta^2 = 0.68$ ; Step 2,  $F(1, 11) = 18.97, p < .002, \eta^2 = 0.63$ ; Step 3,  $F(1, 11) = 13.66, p < .004, \eta^2 = 0.55$ ; Step 4,  $F(1, 11) = 11.62, p < .006, \eta^2 = 0.51$ ; and Step 5,  $F(1, 11) = 19.68, p < .002, \eta^2 = 0.64$ . At the disyllabic ends, however, the fall continuum yielded a reliably lower proportion of two-beat responses relative to the rise continuum,  $F(1, 11) = 15.37, p < .003, \eta^2 = 0.58$ , a result found also with speech controls.

Overall, these results could demonstrate that Russian speakers extended their linguistic knowledge to nonspeech inputs. Recall, however, that about half of the Russian participants described the stimuli using some linguistic terms. To address the possibility that the overall pattern of results might be due to participants who perceived the input as speech, we compared the results of participants who identified the input as speechlike and those of participants who did not by means of a 2 (speech response: whether the input was identified as speech or nonspeech)  $\times$  2 (continuum type)  $\times$  6 (vowel duration) ANOVA. None of the effects involving the speech-response factor approached significance (all  $ps > .33$ ). Another possibility is that participants learned to recognize the stimuli as speech throughout the multiple blocks of the experimental session. However, the pattern of results during the first encounter with each stimulus, at the first experimental block, was similar to the findings across blocks (see Table 2). Russian participants tended to identify the items as having one beat, and the proportion of one-beat responses was overall higher for the rise continuum. A 2 (continuum type)  $\times$  6 (vowel duration) ANOVA over responses to the first block yielded a reliable interaction,  $F(5, 45) = 2.70, p < .04, \eta^2 = 0.23$ . The simple effect of continuum type was significant in Step 2,  $F(1, 11) = 11.00, p < .007, \eta^2 = 0.50$ .

To determine whether responses to nonspeech were modulated by linguistic experience, we next compared the Russian participants with the group of English speakers from Experiment 2 by means of a 2 (language)  $\times$  2 (continuum type)  $\times$  6 (vowel duration) ANOVA (the means are presented in Figure 6). Language experience modulated the effect of vowel duration,  $F(5, 110) = 9.84, p < .0001, \eta^2 = 0.31$ , and these two factors further interacted with continuum type,  $F(5, 110) = 5.56, p < .0002, \eta^2 = 0.20$ . To interpret this interaction, we compared the two groups with respect to the rise and fall continua, separately.

A 2 (language)  $\times$  6 (vowel duration) ANOVA conducted on the rise continuum did not yield a main effect of language or a

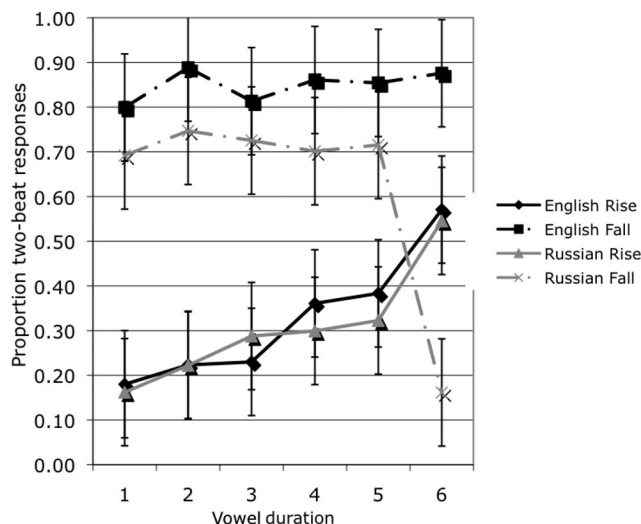


Figure 6. The proportion of two-beat responses to nonspeech stimuli by Russian speakers (in Experiment 5) and English participants (in Experiment 2). Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

Language  $\times$  Vowel Duration interaction (both  $ps > .8$ ), suggesting that language did not modulate responses to the rise continuum. A similar analysis of the fall continuum did yield a significant interaction,  $F(5, 110) = 20.09, p < .0001, \eta^2 = 0.48$ , but this effect was entirely due to the disyllabic end of the continuum. Tukey HSD contrasts showed that the two groups did not differ in their responses to Steps 1–5 (all  $ps > .14$ ), but they did differ significantly on Step 6 ( $p < .0002$ ).

To summarize, like their English counterparts, Russian speakers were sensitive to the structure of nonspeech stimuli. Both groups were more likely to identify onsets of falling sonority as disyllabic. Although nonspeech stimuli attenuated some of the group differences observed with the fall continuum using speech controls, the responses of the two groups to nonspeech stimuli clearly diverged, and this effect was evident in the difficulty of Russian speakers to identify the disyllabic end points as having two beats. Because these end points were unspliced, the lack of discontinuity incorrectly promoted monosyllabic responses. This pattern, seen also with English speakers, suggests that the unfamiliarity with the synthetic inputs encourages attention to acoustic detail. Unlike English, however, Russian lacks  $C\alpha CVC$ , so the unfamiliarity of Russian speakers with synthetic  $C\alpha CVC$  inputs is even greater, and consequently, they are more vulnerable to these misleading acoustic cues. Although these effects of splicing were unexpected, they reflect the effect of linguistic experience on the identification of nonspeech inputs. As a whole, the results suggest that speakers extend systematic knowledge to the processing of nonspeech stimuli, including both principles that are possibly universal and language-particular knowledge.

### Part 3: The Effect of the Speech Manipulation on English Versus Russian Speakers

What knowledge guides the identification of the rise and fall continua used in our experiments? One possibility is that partici-

Table 2  
Proportion of Two-Beat Responses by Russian Speakers in the First Block of Experiment 5

Vowel duration	Continuum type	
	Rise	Fall
Step 1	0.10	0.30
Step 2	0.20	0.70
Step 3	0.10	0.30
Step 4	0.40	0.50
Step 5	0.70	0.40
Step 6	0.20	0.40

T2

F6

pants consult linguistic knowledge—either phonological or phonetic. The similarity in responses to speech and nonspeech stimuli would thus suggest that linguistic knowledge constrains the identification of nonspeech stimuli. An alternative auditory explanation, however, is that participants rely only on general auditory strategies. Although, at first blush, this possibility appears incompatible with the marked effect of linguistic experience on the identification of nonspeech stimuli, several authors have argued that linguistic experience modulates general auditory preferences (e.g., Iversen & Patel, 2008; Iverson et al., 2003; Patel, 2008; Wong, Skoe, Russo, Dees, & Kraus, 2007). Adapting this view to the present results, it is possible that English and Russian participants performed the beat-count task by relying on generic auditory mechanisms. Such mechanisms might lead both groups to attend to the same set of acoustic cues, but because Russian speakers have had greater experience in processing those cues, they were more sensitive to their presence than were English participants. The effect of linguistic experience on processing nonspeech stimuli indicates only different levels of auditory sensitivity, rather than the engagement of the language system.

Our results allow us to evaluate this explanation by comparing the responses of the two groups of participants to speech, speech-control, and nonspeech stimuli. Recall that this spectrum was generated by progressively removing the higher speech formants from natural speech: Speech controls had severe attenuation of these frequencies, whereas in the nonspeech materials, they were completely eliminated. The comparison of natural speech with speech-control and nonspeech stimuli can thus gauge the reliance of the two groups on this acoustic information. If performance is determined by a single set of cues, then the removal of these cues should affect both groups in a qualitatively similar fashion. For example, if cues in the higher frequency range inform the identification of the pretonic vowel, then their removal will reduce the rate of two-beat responses for both groups, and this effect will be mostly evident along the disyllabic end of the vowel-continuum. In contrast, if the two groups rely on different kinds of representations—either because participants attend to different acoustic cues or because they consult different linguistic knowledge—then the removal of a single set of acoustic cues could potentially have qualitatively distinct consequences on the performance of English and Russian participants.

To adjudicate between these views, we compared the responses of the two groups to the natural-speech, speech-control, and nonspeech stimuli by means of a 2 (language) × 3 (speech status: natural speech, first formant control, and nonspeech) × 2 (continuum type: rise vs. fall) × 6 (vowel duration) ANOVA. Given the previous results, we would expect the perception of the rise and fall continua to be modulated by vowel duration and, possibly, speech status. But crucially, if both groups used these cues in a similar fashion, then this interaction should not be further modulated by linguistic experience. The omnibus ANOVA, however, yielded a highly significant four-way interaction,  $F(10, 330) = 5.48, p < .0001, \eta^2 = 0.14$ . Although this interaction did not reach significance in the logit model (see Appendix), and as such, its interpretation requires caution, subsequent comparisons of the effect of speech status separately for English and Russian speakers suggested that speech status had a profoundly different effect on the two groups.

### The Effect of Speech Status on English Speakers

Consider first the effect of speech status on English-speaking participants (see Figure 7). An inspection of the means suggests that the effect of speech status is strongest along the disyllabic end of the vowel continuum: English speakers were more likely to identify disyllabic inputs as having two beats given natural speech compared with synthetic speech and nonspeech stimuli. Moreover, this effect appeared stronger for the rise compared with the fall continuum. A 3 (speech status: natural speech, first formant, and nonspeech) × 2 (continuum type: sonority rise vs. sonority fall) × 6 (vowel duration) ANOVA indeed demonstrated that speech status reliably interacted with continuum type,  $F(2, 33) = 3.81, p < .04, \eta^2 = 0.19$ , and vowel duration,  $F(10, 165) = 5.91, p < .0001, \eta^2 = 0.26$ . The three-way interaction was not significant ( $p > .14$ ).

Analyses of the simple main effects confirmed that speech status reliably modulated responses to the rise continuum,  $F(2, 33) = 6.74, p < .004, \eta^2 = 0.42$ , but not the fall continuum,  $F(2, 33) < 1, \eta^2 = 0.49$  (see Table 3). Specifically, given the rise continuum, natural-speech stimuli were significantly more likely to elicit two-beat responses compared with either speech-control (Tukey HSD tests,  $p < .02$ ) or nonspeech stimuli (Tukey HSD tests,  $p < .0007$ ), which, in turn, did not differ (Tukey HSD tests,  $p > .05$ ).

The effect of speech status on English speakers was also modulated by vowel duration (see Table 4). The simple main effect of speech status was indeed reliable only toward the disyllabic end of the vowel continuum—for Step 6,  $F(2, 33) = 7.55, p < .003, \eta^2 = 0.41$ ; Step 5,  $F(2, 33) = 7.64, p < .002, \eta^2 = 0.41$ ; Step 4,  $F(2, 33) = 8.65, p < .002, \eta^2 = 0.40$ ; Step 3,  $F(2, 33) = 6.16, p < .006, \eta^2 = 0.42$ ; and Step 2,  $F(2, 33) = 1.01, p < .38, \eta^2 = 0.49$ —but not at the monosyllabic end point,  $F(2, 33) < 1, \eta^2 = 0.49$ . Tukey HSD tests showed that two-beat responses were more frequent to speech compared with nonspeech at Steps 3–6 (all  $ps < .003$ ).

### The Effect of Speech Status on Russian Participants

The previous section suggests that, as the acoustic stimulus approximates speech, English speakers were more likely to iden-

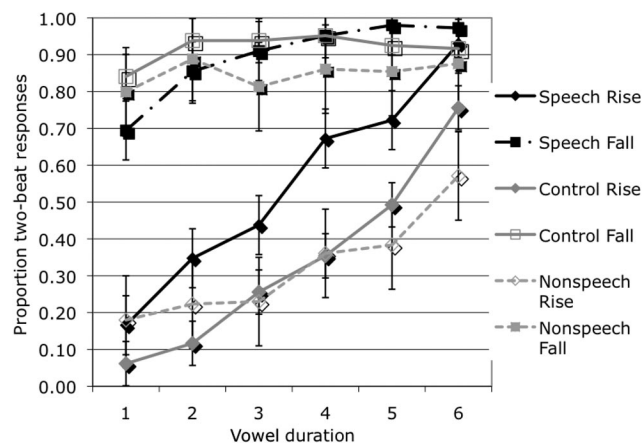


Figure 7. The effect of speech status on English speakers. Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

tify the input as having two beats, and this effect was stronger toward the disyllabic end point of the rise continuum. If this pattern is primarily due to auditory mechanisms that interpret high-formant cues as evidence for two-beat responses, then Russian speakers should exhibit a similar pattern.

A 3 (speech status: natural speech, first formant control, and nonspeech) × 2 (continuum type) × 6 (vowel duration) ANOVA indeed yielded a reliable three-way interaction,  $F(10, 165) = 6.01$ ,  $p < .0001$ ,  $\eta^2 = 0.27$ . But the effect of speech status on Russian participants was markedly different from that of their English-speaking counterparts (see Figure 8). Unlike English speakers, Russian participants were less likely to identify speechlike stimuli as disyllabic, especially toward the monosyllabic end of the fall continuum. The one notable exception to this generalization concerns the unspliced disyllabic end points. Recall that Russian speakers were unable to identify the unspliced synthesized end points (both speech controls and nonspeech) as disyllabic. Accordingly, we examined the effect of speech status separately for the initial five steps.

A 3 (speech status) × 2 (continuum type) × 5 (vowel duration) ANOVA yielded a significant interaction of Speech Status × Vowel Duration,  $F(8, 132) = 4.62$ ,  $p < .0001$ ,  $\eta^2 = 0.22$ , and a marginally significant interaction of Speech Status × Continuum Type,  $F(2, 33) = 3.12$ ,  $p < .06$ ,  $\eta^2 = 0.16$  (see Table 5). Although the interpretation of this finding requires caution, because the interaction was not significant in the logit model (see Appendix), the responses of Russian speakers to the speech manipulation clearly differed from those of English participants. Simple main effect analyses indeed indicated that speech status modulated responses along each of the five steps: Step 1,  $F(2, 33) = 16.08$ ,  $p < .0001$ ,  $\eta^2 = 0.49$ ; Step 2,  $F(2, 33) = 20.51$ ,  $p < .0002$ ,  $\eta^2 = 0.55$ ; Step 3,  $F(2, 33) = 13.53$ ,  $p < .0002$ ,  $\eta^2 = 0.45$ ; Step 4,  $F(2, 33) = 7.67$ ,  $p < .002$ ,  $\eta^2 = 0.32$ ; and Step 5,  $F(2, 33) = 19.85$ ,  $p < .0002$ ,  $\eta^2 = 0.55$ . But unlike English speakers, who tended to identify speech stimuli as disyllabic, Russian participants tended to identify the same stimuli as monosyllabic, and the rate of one-beat responses was significantly higher than that of both speech controls and nonspeech at each of the five steps along the continuum (all  $ps < .05$ ). Similarly, compared with nonspeech stimuli, speech controls were more likely to elicit one-beat responses along each of the three initial steps (all  $ps < .05$ ).

Russian speakers likewise differed from English participants with respect to their sensitivity to the two continua. Recall that for English speakers, speech status enhanced disyllabic responses mostly for the rise continuum. Russians, in contrast, were mostly sensitive to the speech status of falls (see Table 6). Indeed, the simple main effect of speech status was significant only for the fall

Table 3  
Effect of Speech Status and Continuum Type on the Proportion of Two-Beat Responses by English Speakers

Speech status	Continuum type	
	Rise	Fall
Natural speech	0.55	0.89
Speech controls	0.34	0.92
Nonspeech	0.34	0.92

Table 4  
Effect of Vowel Duration and Speech Status on the Proportion of Two-Beat Responses by English Speakers

Vowel duration	Speech status		
	Natural speech	Speech controls	Nonspeech
Step 1	0.43	0.45	0.49
Step 2	0.60	0.53	0.56
Step 3	0.67	0.60	0.52
Step 4	0.81	0.65	0.61
Step 5	0.85	0.71	0.62
Step 6	0.95	0.84	0.72

continuum,  $F(2, 33) = 5.70$ ,  $p < .008$ ,  $\eta^2 = 0.26$ , and not for the rise continuum,  $F(2, 33) = 2.99$ ,  $p < .07$ ,  $\eta^2 = 0.15$ . Tukey HSD tests showed that one-beat responses were more likely for natural speech compared with nonspeech ( $p < .006$ ) stimuli. No other contrast was significant.<sup>3</sup>

### Discussion

This series of analyses examined whether the similarity of the acoustic stimulus to speech has distinct effects on English and Russian participants. We reasoned that if these two groups differ only in the degree of their auditory sensitivity, then the removal of high-frequency formants by the speech manipulation should have qualitatively similar effects on the two groups. The results, however, revealed dramatic differences between English and Russian participants. For English speakers, speechlike stimuli were more likely to elicit disyllabic responses, and this effect was strongest at the disyllabic end point of the rise continuum. By contrast, Russian speakers were more likely to identify speechlike stimuli as monosyllabic, and this effect was stronger for the fall continuum and evident at the monosyllabic end point.

These diametrically opposite responses are inconsistent with the possibility that the two groups differed only on their degree of auditory sensitivity to the same set of acoustic cues. Rather, participants appear to have computed beat count depending on the compatibility of the input with their language-particular knowledge. Indeed, English allows disyllables such as *məlif* and *madif* but disallows the *mlif* or *mdif* type. In contrast, Russian allows nasal-initial onsets of both rising and falling sonority but disallows disyl-

<sup>3</sup> A separate, 3 (speech status: natural speech, speech control, and nonspeech) × 2 (continuum type) ANOVA gauging the effect of speech status for unspliced disyllables at Step 6 yielded a highly significant interaction,  $F(2, 33) = 14.05$ ,  $p < .0002$ ,  $\eta^2 = 0.46$ . Simple main effects analysis demonstrated that speech status significantly modulated responses to both the rise,  $F(2, 33) = 28.36$ ,  $p < .0002$ ,  $\eta^2 = 0.63$ , and fall,  $F(2, 33) = 97.21$ ,  $p < .0002$ ,  $\eta^2 = 0.85$ , continua. Tukey HSD tests suggested that two-beat responses were significantly more frequent given natural-speech compared with either speech-control or nonspeech stimuli, and this effect obtained for both the rise and fall continua (all  $ps < .0002$ ). However, the two types of continua yielded different responses to speech-control and nonspeech stimuli: In the rise continuum, speech controls were reliably less likely to elicit two-beat responses than were nonspeech stimuli ( $p < .0002$ ), but they did not differ given items along the fall continuum ( $p > .90$ ).

F8

T5

Fn3

T6

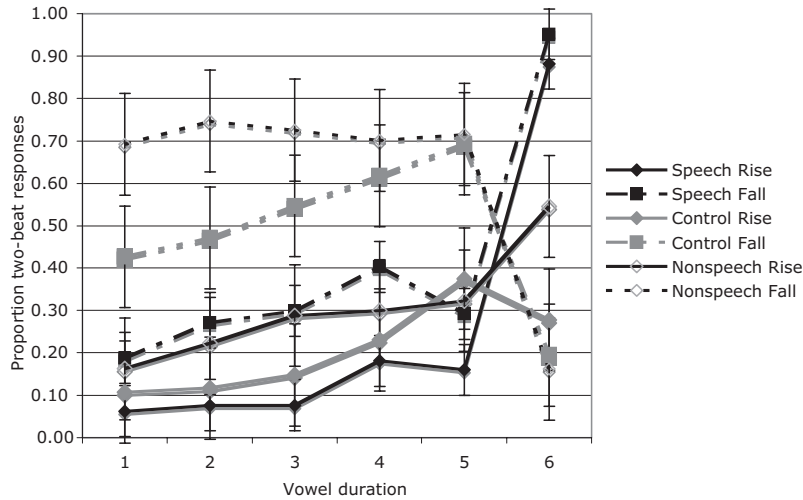


Figure 8. The effect of speech status on Russian speakers. Error bars indicate confidence intervals for the difference between the mean responses to sonority rises and falls.

lables such as *məlif* and *mədif*. The hypothesis that speechlike stimuli activate language-particular knowledge correctly predicts that, for English speakers, speechlike stimuli elicited two-beat responses along the disyllabic end points, whereas Russian speakers were more likely to yield one-beat responses to the monosyllabic ends. While these results do not fully specify the nature of this knowledge or counter all acoustic explanations (we revisit these questions in the General Discussion section), the findings clearly show that the group differences are profound and that they are inexplicable as differences in auditory sensitivity.

**General Discussion**

The nature of the language system is the subject of an ongoing debate. At the center of debate is the specialization of the language system: Are people equipped with mechanisms that are dedicated to the computation of linguistic structure, or is language processing achieved only by domain-general systems? Language universals present an important test for specialization, because their existence could potentially demonstrate structural constraints that are inherent to the language system. To gauge the specialization of the system, however, it is important to determine whether such constraints narrowly apply to both linguistic outputs and inputs.

Table 5  
Effect of Vowel Duration and Speech Status on the Proportion of Two-Beat Responses by Russian Speakers

Vowel duration	Speech status		
	Natural speech	Speech controls	Nonspeech
Step 1	0.13	0.27	0.43
Step 2	0.17	0.29	0.48
Step 3	0.19	0.35	0.51
Step 4	0.29	0.42	0.50
Step 5	0.23	0.53	0.52

The present research examined whether a putatively universal constraint on sonority applies selectively to speech input. Across languages, onsets of falling sonority are systematically dispreferred to onsets of rising sonority (Berent et al., 2007; Greenberg, 1978), and previous research has shown that speakers of various languages extend these preferences even to onsets that are unattested in their language (Berent, 2008; Berent et al., 2010, 2008, 2009, 2007). Sonority preferences trigger the recoding of monosyllables with ill-formed onsets as disyllables (e.g., *lba* → *leba*). Consequently, monosyllables with ill-formed onsets are systematically misidentified as disyllables, and the rate of misidentification is monotonically related to the sonority profile.

The results presented here replicate those previous findings with synthetic materials that were identified by most participants as speech: English speakers were more likely to misidentify monosyllables as having two beats (i.e., as disyllabic) when they included an ill-formed onset of falling sonority (e.g., *mdif*) compared with better formed onsets with a sonority rise (e.g., *mlif*). But surprisingly, the same general pattern emerged even for inputs that did not sound like speech. To determine whether the tendency of English participants to systematically misidentify sonority falls might be due to some generic acoustic properties of the materials, we compared their responses with those of a group of Russian participants, whose language allows nasal clusters of both types. The results showed that Russian speakers could perceive the

Table 6  
Proportion of Two-Beat Responses by Russian Speakers in Steps 1–5

Continuum type	Speech status		
	Natural speech	Speech controls	Nonspeech
Rise	0.11	0.20	0.26
Fall	0.29	0.55	0.72

speechlike stimuli accurately, but their performance differed from that of English participants even when presented with the same nonspeech items. These findings demonstrate that the processing of nonspeech stimuli is modulated by systematic knowledge, including knowledge of principles that are both putatively universal and language-particular.

Taken at face value, the possibility that speech and nonspeech processing is shaped by some common principles does not necessarily challenge the view of the language system as narrowly tuned to linguistic inputs. After all, speech and nonspeech inputs are both auditory stimuli, so it is conceivable that their processing is determined by common sensory constraints. If generic auditory mechanisms could capture the effect of linguistic knowledge seen in our experiments, then the extension of this knowledge to nonspeech stimuli would have no bearing on the input-specificity of the language system.

We consider two versions of the auditory explanation. One possibility is that the observed differences between English and Russian speakers reflect only differences in generic auditory sensitivity. Because Russian speakers have had much more experience in processing sonority falls, they might become more sensitive to the relevant acoustic cues than would English speakers. The detailed comparison of the two groups (in Part 3), however, counters this possibility. An auditory-sensitivity explanation crucially hinges on the assumption that the two groups rely on the same cues—differences are due only to their sensitivity to those cues. But the findings suggest that participants in the two groups relied on different acoustic attributes. The critical findings come from the effect of the speech manipulation—a manipulation that progressively removes high-frequency formants from the speech signal. We reasoned that if the two groups use the acoustic cues present in those high-frequency formants in a similar fashion, then the removal of such cues should affect both groups alike. Contrary to expectations, however, the speech manipulation had profoundly different effects on the two groups. For English speakers, “speechness” affected responses mostly at the disyllabic end of the rise continuum, such that speechlike stimuli were more likely to elicit disyllabic responses. Russian speakers showed the opposite pattern—they were sensitive to speech status mostly for sonority falls. This effect was found already at the monosyllabic end point, and, most critically, speechlike stimuli were more likely to elicit monosyllabic responses. These findings are inconsistent with the possibility that the two groups merely differed in their sensitivity to the high-frequency formants that distinguish speech and nonspeech materials. Clearly, the differences are qualitative and profound.

A second acoustic explanation might concede that linguistic experience can qualitatively alter what acoustic cues inform one’s linguistic responses. There is indeed evidence that speech processing is shaped by minute episodic properties of specific acoustic tokens stored in listeners’ memory (Goldinger, 1996, 1998), and given that English and Russian present their speakers with very different sets of acoustic exemplars, it is conceivable that this experience would alter their strategies of auditory processing. Assuming further that speechlike stimuli activate stored acoustic tokens better than do nonspeech stimuli, this account could potentially explain why speech status increased disyllabic responses for English speakers (i.e., via the activation of similar stored disyllables), whereas for Russian speakers, speechlike stimuli enhanced

monosyllabic responses (i.e., via the activation of similar stored monosyllables).

This auditory account nonetheless faces several challenges. According to this view, performance depends on the acoustic similarity of the input to acoustic tokens encountered in one’s linguistic experience. The finding that Russian speakers were more likely to give one-beat responses to speechlike monosyllables would suggest that such items resemble the acoustic properties of Russian monosyllables more than they resemble English monosyllables. But this assumption is uncertain. Because the experimental stimuli were produced by an English talker, it is not at all evident that they are particularly similar to typical Russian monosyllables. For example, considering only the acoustic level, a spliced English rendition of *melif* may not necessarily resemble an acoustic token of a Russian *ml* monosyllable (e.g., /mlat/, meaning “young”) more than it resembles an English monosyllabic token, say, of *bluff* or *bleak*.

The problem is further exacerbated by the responses of Russian participants. Recall that speech status affected Russian speakers only along the fall continuum but not along the rise continuum. To capture this finding, an acoustic account must assume that *md*-type items resembled Russian monosyllabic tokens more than did *ml*-type tokens. But the actual pattern is opposite: Although Russian has many onsets of falling sonority, the particular onset *md* happens to be absent in Russian, whereas the onset *ml* is frequent (e.g., /mlat/“young”; /mlatʃaj/“junior”; /mlɛtʃnəj/“milky”; /mlɛt/“to be delighted”). In view of the unfamiliarity of Russian participants with the *md* cluster, and the foreign phonetic properties of English speech, it is doubtful that the acoustic properties of the experimental materials are more familiar to Russian participants than to English speakers.

A broader challenge to auditory accounts is presented by the phonetic literature demonstrating that the interpretation of acoustic cues is strongly modulated by linguistic principles rather than generic auditory or motor preferences alone (e.g., Anderson & Lightfoot, 2002; Keating, 1985). In the case of sonority restrictions, there is direct evidence suggesting that the relevant linguistic knowledge is specifically phonological. Nonphonological (auditory and phonetic) explanations assume that the systematic misidentification of onsets of falling sonority reflects an inability to extract the surface form from the acoustic stimulus. Although our present results do not speak to this question, other findings indicate that such difficulties are neither necessary nor sufficient to elicit misidentification. First, when attention to phonetic form is encouraged, English speakers are demonstrably capable of encoding sonority falls accurately, as accurately as they encode better formed onsets (Berent et al., 2010, 2007). Moreover, the misidentification of onsets of falling sonority extends to printed inputs (Berent & Lennertz, 2010; Berent et al., 2009). In fact, sonority effects have been observed for even sign languages (Corina, 1990; Perlmutter, 1992; Sandler, 1993). The generality of sonority effects with respect to stimulus modalities is inconsistent with a purely nonphonological locus—either phonetic or auditory.

Given the numerous challenges to acoustic explanations, one might consider the possibility that the effects of sonority do, in fact, reflect linguistic phonological knowledge. This proposal would nicely account for the distinct effects of speech status on English and Russian speakers, and it easily accommodates the robustness of sonority effects with respect to stimulus modality.

But if phonological principles extend to nonspeech stimuli, then this explanation would force one to abandon the view of the language system as narrowly tuned with respect to its inputs.

Although this conclusion might initially seem inconsistent with the hypothesis of a domain-specific system, on closer scrutiny, it might become all but inevitable. Indeed, language is not restricted to spoken inputs. Not only are sign languages commonly attested, but they are known to emerge spontaneously in Deaf communities and share common structural principles with spoken languages (Sandler, Aronoff, Meir, & Padden, in press; Sandler, Meir, Padden, & Aronoff, 2005; Senghas & Coppola, 2001; Senghas, Kita, & Ozyurek, 2004). In view of such broad language universals, the flexibility of the language system with respect to its inputs is only expected. Such flexibility, however, does not necessarily render the system unconstrained with respect to its input. Rather than narrowly selecting the inputs by some defining physical properties, the viability of a stimulus as linguistic input might depend on the grammatical computations it affords. In this view, the language system might take a broad range of inputs—both auditory and visual—and attempt to analyze them according to a set of domain-specific grammatical constraints. The admissibility of the input might depend on the structural coherence of the output: The better formed the output, the more likely the input is to engage the language system. The proposal that linguistic inputs are identified by their adherence to well-formedness restrictions does not preclude the contribution of articulatory (Liberman & Mattingly, 1989) and auditory (Diehl, Lotto, & Holt, 2004) constraints, nor is it inconsistent with the possibility that speakers also encode indexical properties of the input (Goldinger, 1996, 1998). This proposal would nonetheless suggest that nonspeech stimuli corresponding to well-formed inputs might in fact be judged as more speechlike than their ill-formed counterparts. The higher tendency of Russian participants to spontaneously identify nonspeech stimuli (stimuli corresponding to monosyllables that are well formed in their language) as speechlike and their greater sensitivity to the speech status of monosyllables (well formed in their language) compared with the (worse formed) disyllables are both consistent with this hypothesis.

The possibility that the language system does not strictly constrain its input also carries some practical repercussions. A large neuroimaging literature has attempted to localize the language network by comparing brain activation by speech inputs and various nonspeech controls (e.g., Dehaene-Lambertz et al., 2005; Schofield et al., 2009; Scott et al., 2000; Vouloumanos et al., 2001). Underlying this approach is the assumption that the language system is selectively engaged by speech. But the previous suggestion that stimuli identified as nonspeech nonetheless undergo phonetic processing (Azadpour & Balaban, 2008), along with our current findings, indicate that the linguistic processing of such materials might also engage phonological principles, both universal and language-particular. Together, these results indicate that the use of nonspeech controls might systematically underestimate the identification of the language network (Azadpour & Balaban, 2008). Whether the principles guiding this network are specific to language remains the subject of an ongoing debate, but its inputs are clearly not limited to stimuli that are consciously identified as linguistic.

## References

- Anderson, S. R., & Lightfoot, D. W. (2002). *The language organ: Linguistics as cognitive physiology*. Cambridge, England: Cambridge University Press.
- Azadpour, M., & Balaban, E. (2008). Phonological representations are unconsciously used when processing complex, non-speech signals. *PLoS ONE*, *3*, e1966.
- Berent, I. (2008). Are phonological representations of printed and spoken language isomorphic? Evidence from the restrictions on unattested onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1288–1304.
- Berent, I., & Lennertz, T. (2010). Universal constraints on the sound structure of language: Phonological or acoustic? *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 212–223.
- Berent, I., Lennertz, T., & Balaban, E. (2010). *Language universals and misidentification: A two-way street*. Manuscript submitted for publication.
- Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences, USA*, *105*, 5321–5325.
- Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology*, *26*, 75–108.
- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591–630.
- Best, C. T., Morriongiello, B. A., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191–211.
- Blevins, J. (2004). *Evolutionary phonology*. Cambridge, England: Cambridge University Press.
- Bybee, J. L. (2008). Formal universals as emergent phenomena: The origins of structure preservation. In J. Good (Ed.), *Linguistic universals and language change* (pp. 108–121). New York, NY: Oxford University Press.
- Chomsky, N. (1980). *Rules and representations*. New York, NY: Columbia University Press.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 282–333). Cambridge, England: Cambridge University Press.
- Corina, D. P. (1990). Reassessing the role of sonority in syllable structure: Evidence from visual gestural language. In M. Ziolkowski, M. Noske, & K. Deaton (Eds.), *Papers from the 26th Annual Regional Meeting of the Chicago Linguistic Society* (Vol. 2, pp. 33–43). Chicago, IL: University of Chicago.
- Davidson, L. (2006a). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*, 104–137.
- Davidson, L. (2006b). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science*, *30*, 837–862.
- Dehaene-Lambertz, G., Pallier, C., Semiclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*, 21–33.
- de Lacy, P. (2006). *Markedness: Reduction and preservation in phonology*. New York, NY: Cambridge University Press.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578.
- Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for

- prelexical phonological processing in word recognition. *Language and Cognitive Processes*, 5, 491–505.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
- Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429–492.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Greenberg, J. H. (1978). Some generalizations concerning initial and final consonant clusters. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. 2, pp. 243–279). Stanford, CA: Stanford University Press.
- Hallé, P., Segui, J., Frauenfelder, U. H., & Meunier, C. (1998). The processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, 24, 592–608.
- Hayes, B., Kirchner, R. M., & Steriade, D. (Eds.). (2004). *Phonetically based phonology*. Cambridge, England: Cambridge University Press.
- Hooper, J. B. (1976). *An introduction to natural generative phonology*. New York, NY: Academic Press.
- Iversen, J. R., & Patel, A. D. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124, 2263–2271.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–57.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115–132). Orlando, FL: Academic.
- Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10, 421–442.
- Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*. New York, NY: Oxford University Press.
- Kuhl, P. K., & Meltzoff, A. N. (1982, December 10). The bimodal perception of speech in infancy. *Science*, 218, 1138–1141.
- Liberman, A. M., & Mattingly, I. G. (1989, January 27). A specialization for speech perception. *Science*, 243, 489–494.
- Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience*, 15, 549–558.
- Massaro, D. W., & Cohen, M. M. (1983). Phonological constraints in speech perception. *Perception & Psychophysics*, 34, 338–348.
- Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 415–420). Amsterdam, The Netherlands: North Holland.
- Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 2, 131–157.
- McCarthy, J. (2005). Optimal paradigms. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in phonological theory* (pp. 295–371). Oxford, England: Oxford University Press.
- McClelland, J. L. (2009). Phonology and perception: A cognitive scientist's perspective. In P. Boersma & S. Hamann (Eds.), *Phonology in perception*. Berlin, Germany: Mouton De Gruyter.
- Meyer, M., Zaehle, T., Gountouna, V.-E., Barron, A., Jancke, L., & Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *NeuroReport*, 16, 1985–1989.
- Ohala, J. J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers From the 26th Annual Regional Meeting of the Chicago Linguistic Society* (Vol. 2, pp. 319–338). Chicago, IL: University of Chicago.
- Parker, S. (2002). *Quantifying the sonority hierarchy*. Unpublished doctoral dissertation, Department of Linguistics, University of Massachusetts at Amherst.
- Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.
- Peperkamp, S. (2007). Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition*, 104, 638–643.
- Perlmutter, D. M. (1992). Sonority and syllable structure in American Sign Language. *Linguistic Inquiry*, 407–442.
- Pinker, S. (1994). *The language instinct*. New York, NY: Morrow.
- Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. Christiansen & S. Kirby (Eds.), *Language evolution: States of the art*. New York, NY: Oxford University Press.
- Pinker, S., & Bloom, P. (1994). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707–784.
- Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, 60, 941–951.
- Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to Universal Grammar. *Science*, 275, 1604–1610.
- Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. (Original work published 1993)
- Redford, M. A. (2008). Production constraints on learning novel onset phonotactics. *Cognition*, 107, 785–816.
- Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science*, 12, 24–29.
- Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, 101, 129–156.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981, May 22). Speech perception without traditional speech cues. *Science*, 212, 947–949.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning past tense of English verbs: Implicit rules or parallel distributed processing? In J. L. McClelland, D. E. Rumelhart, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 216–271). Cambridge MA: MIT Press.
- Sandler, W. (1993). A sonority cycle in American Sign Language. *Phonology*, 10, 242–279.
- Sandler, W., Aronoff, M., Meir, I., & Padden, C. (in press). The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*.
- Sandler, W., & Lillo-Martin, D. C. (2006). *Sign language and linguistic universals*. Cambridge, England: Cambridge University Press.
- Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences, USA*, 102, 2661–2665.
- Saussure, F. d. (1959). *Course in general linguistics*. New York, NY: Philosophical Library. (Original work published 1915)
- Schofield, T. M., Iverson, P., Kiebel, S. J., Stephan, K. E., Kilner, J. M., Friston, K. J., . . . Leff, A. P. (2009). Changing meaning causes coupling



- changes within higher levels of the cortical hierarchy. *Proceedings of the National Academy of Sciences, USA*, 106, 11765–11770.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123, 2400–2406.
- Selkirk, E. O. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oerhle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students* (pp. 107–136). Cambridge, MA: MIT Press.
- Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12, 323–328.
- Senghas, A., Kita, S., & Ozyurek, A. (2004, September 17). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305, 1779–1782.
- Smolensky, P. (2006). Optimality in phonology II: Markedness, feature domains, and local constraint conjunction. In P. Smolensky & G. Legendre (Eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar: Vol. 2. Linguistic and philosophical implications* (pp. 27–160). Cambridge, MA: MIT Press.
- Smolensky, P., & Legendre, G. (2006). Harmony optimization and the computational architecture of the mind/brain. In P. Smolensky & G. Legendre (Eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar: Vol. 1. Cognitive architecture* (pp. 3–61). Cambridge, MA: MIT Press.
- Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. Unpublished dissertation, Department of Linguistics and Philosophy, Massachusetts Institute of Technology.
- Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition*, 96, B13–22.
- Vennemann, T. (1972). On the theory of syllable phonology. *Linguistische Berichte*, 18, 1–18.
- Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13, 994–1005.
- Warren, R. M., Bashford, J. A. J., & Lenz, P. W. (2004). Intelligibility of bandpass filtered speech: Steepness of slopes required to eliminate transition band contributions. *Journal of the Acoustical Society of America*, 115, 1292–1295.
- Warren, R. M., Bashford, J. A. J., & Lenz, P. W. (2005). Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired. *Journal of the Acoustical Society of America*, 118, 3261–3266.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brain stem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10, 420–422.
- Wright, R. (2004). A review of perceptual cues and robustness. In D. Steriade, R. Kirchner, & B. Hayes (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge, England: Cambridge University Press.
- Zec, D. (2007). The syllable. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 161–194). Cambridge, England: Cambridge University Press.

Appendix

A Multilevel Logit Analysis of Some Key Interactions

Table A1

Experiment/comparison and interaction source	$\beta$	<i>SE</i>	<i>z</i>	<i>p</i>
1				
Continuum Type $\times$ Vowel Duration	-.6884	.09814	-7.01	<.0001
2				
Continuum Type $\times$ Vowel Duration (all trials)	-.3078	.0748	-4.11	<.0001
Continuum Type $\times$ Vowel Duration (first block only)	-.4072	.1774	-2.23	<.03
3				
Continuum Type $\times$ Vowel Duration	-.4405	.0910	-4.84	<.0001
Continuum Type $\times$ Vowel Duration $\times$ Language	-.3603	.1394	-2.58	<.001
Vowel Duration $\times$ Language (sonority rises only)	.2521	.0969	2.60	<.001
Vowel Duration $\times$ Language (sonority falls only)	-.1676	.1055	-1.59	<.12
4				
Continuum Type $\times$ Vowel Duration	-.6596	.3246	-2.03	<.05
Continuum Type $\times$ Vowel Duration $\times$ Language	.2351	.1154	2.04	<.05
Vowel Duration $\times$ Language (sonority falls only)	-.1881	.0859	-2.18	<.03
Vowel Duration $\times$ Language (sonority rises only)	-.4596	.0816	-5.63	<.0001
5				
Continuum Type $\times$ Vowel Duration (all trials)	-.6775	.0646	-10.49	<.0001
Continuum Type $\times$ Vowel Duration (first block only)	-.6433	.1322	-4.87	<.001
Continuum Type $\times$ Vowel Duration $\times$ Language	-.3956	.0972	-4.07	<.0001
Vowel Duration $\times$ Language (sonority falls only)	-.4888	.0756	-6.47	<.0001
Vowel Duration $\times$ Language (sonority rises only)	-.0408	.0672	<1	
Cross-language comparison				
Speech Status $\times$ Continuum Type $\times$ Language	-.1355	.0403	-3.36	<.0001
Speech Status $\times$ Continuum Type $\times$ Vowel Duration $\times$ Language	-.0331	.0806	<1	
The effect of speech status on English speakers (Steps 1–6)				
Speech Status $\times$ Vowel Duration	-.2447	.0314	-7.79	<.0001
Speech Status $\times$ Continuum Type $\times$ Vowel Duration	-.1142	.0627	-1.83	<.07
The effect of speech status on Russian speakers (Steps 1–6)				
Speech Status $\times$ Vowel Duration	-.3707	.0259	-14.26	<.0001
Speech Status $\times$ Vowel Duration $\times$ Continuum Type	-.1507	.0519	-2.90	<.001
The effect of speech status on Russian speakers (Steps 1–5)				
Speech Status $\times$ Vowel Duration	-.0761	.0340	-2.24	<.03
Speech Status $\times$ Continuum Type	-.4016	.0968	4.15	<.0001
Speech Status $\times$ Continuum Type $\times$ Vowel Duration	-.0108	.0681	<1	
The effect of speech status on Russian speakers (Step 6)				
Speech Status $\times$ Continuum Type	-1.5342	.2393	-6.41	<.0001

Received October 23, 2009  
 Revision received March 18, 2010  
 Accepted March 23, 2010 ■