frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 13 September 2011
doi: 10.3389/fpsyg.2011.00182

How linguistic chickens help spot spoken-eggs: phonological constraints on speech identification

Iris Berent[1]*, Evan Balaban[2,3] and Vered Vaknin-Nusbaum[4,5]

[1] *Department of Psychology, Northeastern University, Boston, MA, USA*
[2] *Department of Psychology, McGill University, Montreal, QC, Canada*
[3] *Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy*
[4] *Department of Education, University of Haifa, Haifa, Israel*
[5] *Western Galilee College, Akko, Israel*

Edited by:
Charles Clifton Jr., University of Massachusetts Amherst, USA

Reviewed by:
Chen Yu, Indiana University, USA
Bob McMurray, University of Iowa, USA

****Correspondence:***
Iris Berent, Department of Psychology, Northeastern University, 125 Nightingale Hall, 360 Huntington Avenue, Boston, MA 02115-5000, USA.
e-mail: i.berent@neu.edu

It has long been known that the identification of aural stimuli as speech is context-dependent (Remez et al., 1981). Here, we demonstrate that the discrimination of speech stimuli from their non-speech transforms is further modulated by their linguistic structure. We gauge the effect of phonological structure on discrimination across different manifestations of well-formedness in two distinct languages. One case examines the restrictions on English syllables (e.g., the well-formed *melif* vs. ill-formed *mlif*); another investigates the constraints on Hebrew stems by comparing ill-formed AAB stems (e.g., *TiTuG*) with well-formed ABB and ABC controls (e.g., *GiTuT, MiGuS*). In both cases, non-speech stimuli that conform to well-formed structures are harder to discriminate from speech than stimuli that conform to ill-formed structures. Auxiliary experiments rule out alternative acoustic explanations for this phenomenon. In English, we show that acoustic manipulations that mimic the *mlif–melif* contrast do not impair the classification of non-speech stimuli whose structure is well-formed (i.e., disyllables with phonetically short vs. long tonic vowels). Similarly, non-speech stimuli that are ill-formed in Hebrew present no difficulties to English speakers. Thus, non-speech stimuli are harder to classify only when they are well-formed in the participants' native language. We conclude that the classification of non-speech stimuli is modulated by their linguistic structure: inputs that support well-formed outputs are more readily classified as speech.

Keywords: phonology, speech, non-speech, well-formedness, phonological theory, modularity, encapsulation

INTRODUCTION

Speech is the preferred carrier of linguistic messages. All hearing communities use oral sound as the principal medium of linguistic communication (Maddieson, 2006); from early infancy, people favor speech stimuli to various aural controls (e.g., Vouloumanos and Werker, 2007; Shultz and Vouloumanos, 2010; Vouloumanos et al., 2010); and speech stimuli may engage many so-called language areas in the brain to a greater extent than non-speech inputs (Molfese and Molfese, 1980; Vouloumanos et al., 2001; Liebenthal et al., 2003; Meyer et al., 2005; Telkemeyer et al., 2009; but see Abrams et al., 2010; Rogalsky et al., 2011).

The strong human preference for speech suggests that the language system is highly tuned to speech. This is indeed expected by the view of the language system as an adaptive processor, designed to ensure a rapid automatic processing of linguistic messages (Liberman et al., 1967; Fodor, 1983; Liberman and Mattingly, 1989; Trout, 2003; Pinker and Jackendoff, 2005). But surprisingly, the preferential tuning to speech is highly flexible. And indeed, linguistic phonological computations apply not only to aural language, but also to printed stimuli read silently (e.g., Van Orden et al., 1990; Lukatela et al., 2001, 2004; Berent and Lennertz, 2010). Moreover, many natural languages take manual signs as their inputs, and such inputs spontaneously give rise to phonological systems that mirror several aspects of spoken language phonology (Sandler and Lillo-Martin, 2006; Sandler et al., 2011; Brentari et al., 2011). Finally, the classification of linguistic stimuli as speech can be strategically altered by instructions and practice (e.g., Remez et al., 1981, 2001; Liebenthal et al., 2003; Dehaene-Lambertz et al., 2005).

Such results make it clear that the identification of speech – the linguistic messenger – is logically distinct from the message. Not only can we use speech to transmit messages that lack both meaning and grammatical structure (e.g., *o-i-a*), but it is also possible to convey meaningful, well-formed linguistic messages using acoustic information that can be perceived as non-speech-like (e.g., sine wave analogs). In practice, however, the message and messenger interact, as the classification of an input as "linguistic" can be flexibly altered by numerous factors in a top-down fashion. This flexibility raises a conundrum. An adaptive language system should rapidly hone in on its input and readily discern linguistic messengers from non-linguistic ones. But if the status of an auditory input is not determined solely by its internal acoustic properties, then how do we classify it as "linguistic"? Specifically, what external factors constrain the classification of an input as "speech"?

Most existing research has addressed this question by pitting internal, bottom-up factors against top-down effects that are external to the stimulus (e.g., task demands, training). Top-down effects, however, might also originate from the language system itself. The language system is a computational device that generates detailed structural descriptions to inputs and evaluates their well-formedness (Chomsky, 1980; Prince and Smolensky,

1993/2004; Pinker, 1994). Existing research has shown that such computations might apply to a wide range of inputs – both inputs that are perceived as speech, and those classified as non-speech-like. To the extent the system is interactive, it is thus conceivable that the classification of an input as "linguistic" might be constrained by its output – namely, its structural well-formedness. Such top-down effects could acquire several forms. On a weaker, attention-based explanation, ill-formed stimuli are less likely to engage attentional resources, so they allow for a more rapid and accurate classification of the stimulus, be it speech or non-speech. A stronger interactive view asserts that the output of the computational system can inform the interpretation of its input – the stronger the well-formedness of the output (e.g., harmony, Prince and Smolensky, 1997), the more likely the input is to be interpreted as linguistic. Accordingly, ill-formedness should facilitate the rapid classification of non-speech inputs, but impair the classification of speech stimuli. While these two versions differ on their accounts for the classification of speech inputs, they converge on their predictions for non-speech stimuli: ill-formed inputs will be more readily classified as non-speech compared to well-formed structures.

Past research has shown that the identification of non-speech stimuli is constrained by several aspects of linguistic knowledge. Azadpour and Balaban (2008) observed that people's ability to discriminate non-speech syllables from each other depends on their phonetic distance: the larger the phonetic distance, the more accurate the discrimination. Moreover, this sensitivity to the phonetic similarity of non-speech stimuli remains significant even after statistically controlling for their acoustic similarity (determined by the Euclidian distance among formants).

Subsequent research has shown that the identification of non-speech stimuli is constrained by phonological knowledge as well (Berent et al., 2010). Participants in these experiments were presented with various types of auditory continua – either natural speech stimuli, non-speech stimuli, or speech-like controls – ranging from a monosyllable (e.g., *mlif*) to a disyllable (e.g., *melif*), and they were instructed to identify the number of their "beats" (a proxy for syllables). Results showed that syllable count responses were modulated by the phonological well-formedness of the stimulus, and the effect of well-formedness obtained regardless of whether the stimulus was perceived as speech or non-speech.

These results demonstrate that people can compute phonological structure (a property of linguistic messages) for messengers that they classify as non-speech. But other aspects of the findings suggest that the structure of the message can further shape the classification of messenger. The critical evidence comes from the comparison of speech and non-speech stimuli. As expected, responses to speech and non-speech stimuli differed – a difference we dub the "speechiness" effect. But remarkably, the "speechiness" effect was stronger for well-formed stimuli compared to ill-formed ones. Well-formedness, here, specifically concerned the contrast between monosyllables (e.g., *mlif*) and their disyllabic counterparts (e.g., *melif*) in two languages: English vs. Russian. While English allows *melif*-type disyllables, but not their monosyllabic *mlif*-type counterparts, Russian phonotactics are opposite – Russian allows sequences like *mlif*, but bans disyllables such as *melif* (/məlif/). The experimental results showed that the Russian and English groups were each sensitive to the status of the stimuli as speech or non-speech, but the "speechiness" effect depended on the well-formedness of the stimuli in the participants' language. English speakers manifested a stronger "speechiness" effect for *melif*-type inputs, whereas for Russian speakers, this effect was more robust for monosyllables – structures that are well-formed in their language. Interestingly, this well-formedness effect obtained irrespective of familiarity – for both *mlif*-type items (which are both well-formed and attested in Russian) and the *mdif*-type – items that are structurally well-formed (Russian exhibits a wide range of sonorant-obstruent onsets), but happen to be unattested in this language. These results suggest that the classification of auditory stimuli as speech depends on their linguistic well-formedness: well-formed stimuli are more "speech-like" than ill-formed controls. Put differently, structural properties of the linguistic message inform the classification of the messenger.

The following research directly tests this prediction. Participants in these experiments were presented with a mixture of speech and non-speech stimuli, and they were asked simply to determine whether or not the stimulus sounds like speech. The critical manipulation concerns the well-formedness of those stimuli. Specifically, we compare responses to stimuli generated from inputs that are either phonologically well-formed or ill-formed. Our question here is whether the ease of discriminating speech from non-speech might depend on the phonological well-formedness of these non-speech stimuli. The precise source of this well-formedness effect (whether it is due to a weaker effect of attention-grabbing, or a strong top-down interaction) is a question that we defer to the Section "General Discussion." For this reason, we make no *a priori* predictions regarding the effect of well-formedness on speech stimuli. Our goal here is to first establish that well-formedness modulates the classification of non-speech inputs. To the extent that well-formed stimuli are identified as speech-like, we expect that participants should exhibit consistent difficulty in the classification of non-speech stimuli whose structure is well-formed.

Well-formed stimuli, however, might also raise difficulties for a host of acoustic reasons that are unrelated to phonological structure. Our investigation attempts to distinguish phonological well-formedness from its acoustic correlates in two ways. First, we examine the effect of well-formedness across two different languages, using two manifestations that differ on their phonetic properties – Experiment 1 examines the restrictions on syllable structure in English, whereas Experiment 3 explores the constraints on stem structure in Hebrew. Second, we demonstrate that the effect of well-formedness is dissociable from the acoustic properties of the input. While Experiments 1 and 3 compare well-formed stimuli to ill-formed counterparts, Experiment 2 and Experiment 4 each applies the same phonetic manipulations to stimuli that are phonologically well-formed. Specifically, Experiment 2 shows that a phonetic manipulation comparable to the one used in Experiment 1 fails to produce the same results for stimuli that are well-formed, whereas Experiment 4 demonstrates that the difficulties with non-speech stimuli that are well-formed in Hebrew are eliminated once the same stimuli are presented to English speakers.

PART 1: ENGLISH SYLLABLE STRUCTURE CONSTRAINS THE CLASSIFICATION OF NON-SPEECH STIMULI

Experiments 1–2 examine whether the classification of acoustic stimuli as speech depends on their well-formedness as English syllables. Participants in this experiment were presented with a mixture of non-speech stimuli and matched speech-like controls, and they were simply asked to determine whether or not each stimulus sounds like speech. Of interest is whether the classification of the stimulus as speech depends on its well-formedness.

To examine this question, we simultaneously manipulated both the phonological structure of the stimuli and their speech status. Phonological well-formedness was manipulated along nasal-initial continua ranging from well-formed disyllables (e.g., /məlɪf/, /mə dɪf) to ill-formed monosyllables (e.g., /mlɪf/, /mdɪf/). To generate these continua, we first had a native English talker naturally produce a disyllable that began with a nasal–schwa sequence – either one followed by a liquid (e.g., /məlɪf/) or one followed by a stop (e.g., /mə dɪf/). We then gradually excised the schwa in five steps until, at the last step, the schwa was entirely removed, resulting in a CCVC monosyllable – either /mlɪf/ or /mdɪf/ (we use C and V for consonants and vowels, respectively). The CCVC–CəCVC continuum thus presents a gradual contrast between well-formed inputs (in step 6, CəCVC) and ill-formed ones (in step 1, CCVC). Among these two CCVC monosyllables, mdɪf-type stimuli are worse formed than their mlɪf-type counterparts (Berent et al., 2009). Although past research has shown that English speakers are sensitive to the *ml–md* distinction given these same materials – both speech and non-speech (Berent et al., 2009, 2010, in press), it is unclear whether this subtle contrast can modulate performance in a secondary speech-detection task. Our main interest, however, concerns the contrast between CCVC monosyllables (either *mlif* or *mdif*) and their CəCVC counterparts. Across languages, complex onsets (in the monosyllable *mlif*) are worse formed than simple onsets (in the disyllable *melif*, Prince and Smolensky, 1993/2004). Nasal-initial complex onsets, moreover, are utterly unattested in English. Accordingly, the monosyllables at step 1 of our continuum are clearly ill-formed relative to the disyllabic endpoints. Our question here is whether ill-formedness of those monosyllables would facilitate their classification as non-speech.

To address this question, we next modified those continua to generate non-speech inputs and speech controls. Non-speech stimuli were produced by resynthesizing the first formant of the natural speech stimuli. To assure that differences between non-speech and speech-like stimuli are not artifacts of the re-synthesis process, we compared those non-speech stimuli to more speech-like inputs that were similarly filtered. If well-formed stimuli are more speech-like, then non-speech responses should be harder to make for well-formed non-speech stimuli – those corresponding to the disyllabic items – compared to ill-formed monosyllables. Accordingly, the identification of non-speech stimuli should be modulated by vowel duration. Experiment 1 verifies this prediction; Experiment 2 rules out alternative non-linguistic explanations for these findings.

EXPERIMENT 1

Method

Participants. Ten native English speakers, students at Northeastern University took part in the experiment in partial fulfillment of course requirements.

Materials. The materials included the three pairs of nasal $C_1C_2VC_3$–$C_1əC_2VC_3$ non-speech and speech-control continua used in Berent et al. (2010). Members of the pair were matched for their rhyme and the initial consonant (always an *m*), and contrasted on the second consonant – either *l* or *d* (/mlɪf/-/mdɪf/, /mlɛf/-/mdɛf/, /mlɛb/-/mdɛb/). To generate those continua, we first had an English talker naturally produce the disyllabic counterparts of each pair member (e.g., /məlɪf/, /mədɪf/) and selected disyllables that were matched for length, intensity, and the duration of the pretonic schwa. We next continuously extracted the pretonic vowel at zero crossings in five steady increments, moving from its center outwards. 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 used to generate non-speech stimuli and speech-like stimuli using the procedure detailed in Berent et al. (2010). Briefly, non-speech materials were generated by deriving the first formant contours from spectrograms of the original speech stimuli (256 point DFT, 0.5 ms time increment, Hanning window) using a peak-picking algorithm, which also extracted the corresponding amplitude values. A voltage-controlled oscillator modulated by the amplitude contour was used to resynthesize these contours back into sounds, and the amplitude of the output was adjusted to approximate the original stimulus. The more "speech-like" controls were generated using a digital low-pass filter with a slope of −85 dB per octave above a cutoff frequency that was stimulus-dependent (1216 Hz for /məlɪf/ and /mədɪf/-type items, 1270 Hz for /məlɛf/, 1110 Hz for /məlɛb/, 1347 Hz for /mədɛf/, and 1250 Hz for /mədɛb/-type items), designed to reduce but not eliminate the speech information available at frequencies higher than the cutoff frequency. This manipulation was done as a "control" manipulation to acoustically alter the stimuli in a similar manner to the non-speech stimuli, while preserving enough speech information for these items to be identified as (degraded) speech. Previous testing using these materials confirmed that they were indeed identified as intended (speech or non-speech) by native English participants (Berent et al., 2010). **Figure 1** provides an illustration of the non-speech materials and controls; a sample of the materials is available at http://www.psych.neu.edu/faculty/i.berent/publications.htm.

The six-step continuum for each of the three pairs was presented in all six durations for both non-speech stimuli and speech controls, resulting in a block of 72 trials. Each such block was repeated three times, yielding a total of 216 trials. The order of trials within each block was randomized.

Procedure. Participants were wearing headphones and seated in front of the computer screen. Each trial began with a message indicating the trial number. Participants initiated the trial by pressing the spacebar, which, in turn, triggered the presentation of a fixation point (+, presented for 500 ms) followed by an auditory stimulus. Participants were asked to determine as quickly and accurately as possible whether or not the stimulus corresponded to speech, and indicate their response by pressing

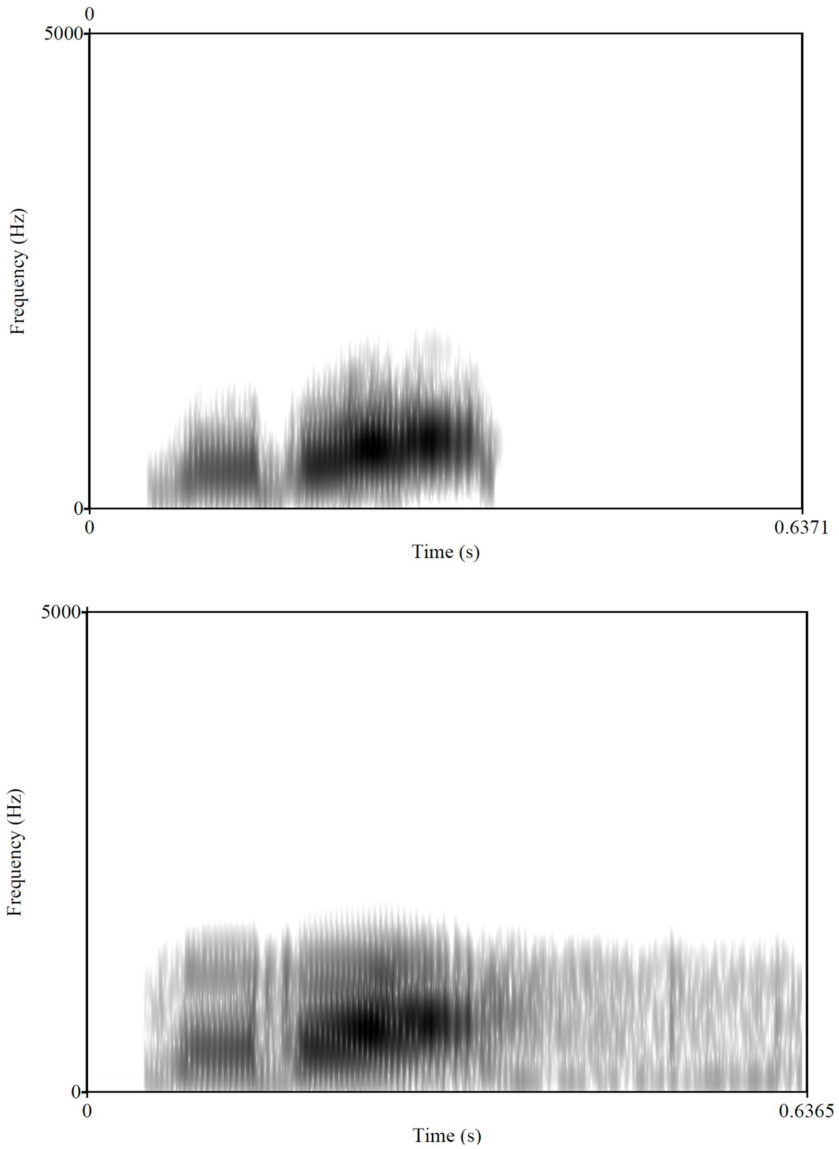


FIGURE 1 | Spectrograms of "*mdif*" (in step 1). Top panel shows the non-speech stimulus; the bottom panel depicts the speech control.

one of two keys on the computer's numeric keypad (1 = speech, 2 = non-speech). Their response was timed relative to the onset of the stimulus. Slow (responses slower than 1000 ms) and inaccurate responses triggered a warning message from the computer. Prior to the experiment, participants received a short practice session with similar items that did not appear in the experimental session.

Results and Discussion

Outliers (correct responses falling 2.5 SD beyond the mean, or faster than 200 ms, less than 3% of the total correct responses) were removed from the analyses of response time. Mean response time and response accuracy are provided in **Table 1**. An inspection of those means confirmed that participants indeed classified the speech and non-speech stimuli as intended ($M = 97\%$).

To determine whether speech and non-speech inputs were affected by syllable structure, we first compared speech and non-speech inputs by means of a 2 speech status × 6 vowel duration × 2 continuum type (*ml* vs. *md*) ANOVA. Because each condition in this experiment includes only three items, these analyses were conducted using only participants as a random variable. The analysis of response accuracy only produced a significant main effect of speech status [$F(1, 9) = 7.12$, MSE = 0.0052, $p < 0.03$], indicating that people responded more accurately to speech-like stimuli compared to their non-speech counterparts. No other effect was significant (all $p > 0.11$).

The analysis of response time, however, yielded a reliable effect of vowel duration [$F(5, 45) = 5.20$, MSE = 978, $p < 0.0008$] as well as a significant three way interaction [$F(5, 45) = 2.42$, MSE = 1089, $p = 0.050$]. No other effect was significant (all $p > 0.13$). We thus

proceeded to investigate the effect of linguistic structure for speech and non-speech stimuli separately by means of 2 continuum type × 6 vowel duration ANOVAs.

Figure 2 plots the effect of vowel duration on speech-like and non-speech stimuli. An inspection of the means suggests that, as the duration of the vowel increased, people took longer to respond to non-speech stimuli. In contrast, response to speech-like stimuli was not monotonically linked to vowel duration.

Table 1 | Mean response accuracy and response time to speech and non-speech stimuli in Experiment 1.

	Vowel duration	Stimulus type			
		Speech		Non-speech	
		ml	*md*	*ml*	*md*
Response accuracy (proportion correct)	1	0.98	0.99	0.94	0.98
	2	0.99	0.97	0.94	0.94
	3	1.00	0.99	0.97	0.94
	4	0.98	0.98	0.97	0.99
	5	0.99	1.00	0.92	0.97
	6	0.98	0.99	1.00	0.96
Response time (ms)	1	588	592	571	566
	2	612	618	587	593
	3	601	608	594	588
	4	608	586	573	622
	5	599	624	589	605
	6	622	607	610	615

The ANOVAs (6 vowel duration × 2 continuum) conducted on speech-like stimuli produced no significant effects in either response time (all $p > 0.15$) or accuracy (all $F < 1$). The linguistic structure of the stimuli also did not reliably affect response accuracy to non-speech inputs (all $p > 0.15$). In contrast, response time to non-speech stimuli was reliably modulated by their linguistic structure. The 6 vowel duration × 2 continuum ANOVA on non-speech stimuli yielded a significant effect of vowel duration [$F(5, 45) = 4.12$, MSE = 979, $p < 0.005$]. Tukey HSD tests revealed that response to fully monosyllabic stimuli (in step 1) were faster than response to disyllabic stimuli (in step 6, $p < 0.001$), and marginally so relative to steps 5 ($p < 0.07$) and 4 ($p < 0.07$). The same ANOVA also yielded marginally significant effects of continuum type [$F(1, 9) = 3.90$, MSE = 897, $p < 0.09$] and a vowel duration × continuum type interaction [$F(5, 45) = 2.24$, MSE = 2070, $p < 0.07$]. Tukey HSD tests indicated that *md*-type continua produced slower responses than their *ml*-type counterparts at step 4 only ($p < 0.009$). This effect, however, did not concern monosyllables in step 1, so it likely reflects the acoustic properties of some of the *md*-items, rather than their phonological structure. Because the silence associated with stop consonants promotes discontinuity in the phonetic signal, the phonetically bifurcate *md*-stimuli might be more readily identified as disyllabic. Such phonetic cues might be particularly salient when the duration of the pretonic vowel is otherwise ambiguous – toward the middle of the vowel continuum. For this reason, middle-continuum *md*-stimuli might be considered as better formed than *ml*-controls.

The main finding of Experiment 1 is that non-speech stimuli are harder to classify when they correspond to well-formed syllables compared to ill-formed ones. Thus, well-formedness impairs the identification of non-speech stimuli.

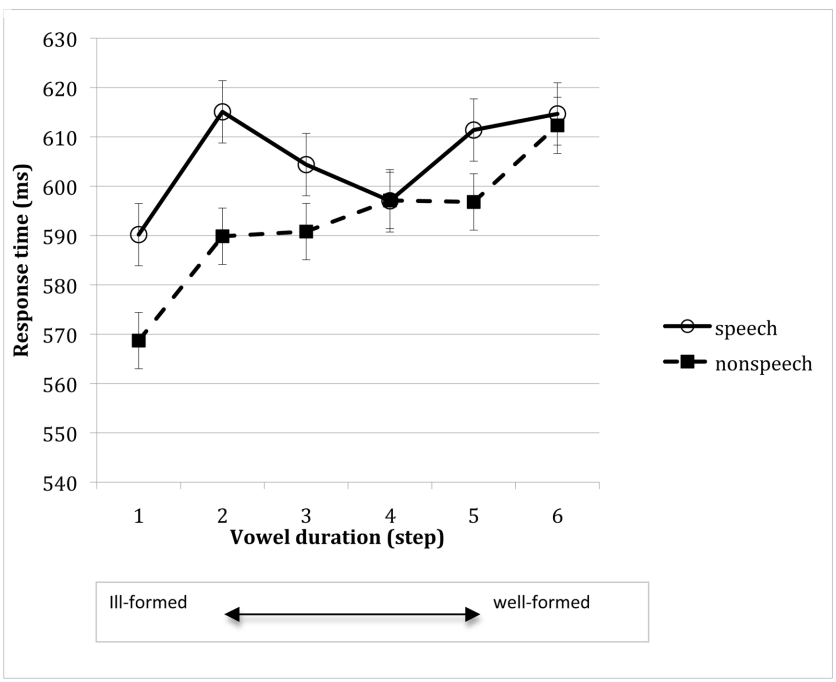


FIGURE 2 | Response time to speech and non-speech stimuli as a function of vowel duration (in Experiment 1). Error bars reflect confidence intervals, constructed for the difference among means along each of the vowel duration continua (i.e., speech and non-speech).

EXPERIMENT 2

The difficulties in responding to well-formed non-speech stimuli could indicate that the classification of non-speech is modulated by phonological well-formedness. Such difficulties, however, could also result from non-linguistic reasons. One concern is that the longer responses to the well-formed disyllables are an artifact of their longer acoustic duration. This explanation, however, is countered by the finding that the very same duration manipulation had no measurable effect on the identification of speech stimuli, so it is clear that response time does not simply mirror the acoustic duration of the stimuli.

Our vowel duration manipulation, however, could have nonetheless affected other attributes of these stimuli that are unrelated to well-formedness. One possibility is that the acoustic cues associated with vowels are more readily identified as speech, so stimuli with longer vowels are inherently more speech-like than short-vowel stimuli. Another explanation attributes the "speechiness" of the disyllabic endpoints to splicing artifacts. Recall that, unlike the other five steps, the sixth endpoint was produced naturally, unspliced. Its greater resemblance to speech could thus result from the absence of splicing.

Experiment 2 addresses these possibilities by dissociating these two acoustic attributes from linguistic well-formedness. To this end, Experiment 2 employs the same vowel manipulation used in Experiment 1, except that the excised vowel was now the tonic (e.g., /ɪ/ in mədɪf), rather than the pretonic vowel (i.e., the schwa). We applied this manipulation to the same naturally produced disyllables used in Experiment 1, and we gradually decreased the vowel duration along the same six continuum-step employed in Experiment 1, such that the difference between steps 1 and 6 in the two experiment was closely matched. This manipulation thus replicates the two acoustic characteristics of the pretonic vowel continua – it gradually decreases the acoustic energy of a vowel, and it contrasts between spliced vowels (in step 1–5) and unspliced ones (in step 6). Unlike Experiment 1, however, this manipulation did not fully eliminate the vowel but only reduced its length, such that the short- and long-endpoints were clearly identified as disyllables. Since the vowel endpoints do not contrast phonologically in English, the increase in the duration of the tonic vowel (in Experiment 2) does not alter the phonological structure of these stimuli.

If the difficulty responding to non-speech stimuli with longer pretonic vowels (in Experiment 1) is due to the acoustic properties of vowels, then non-speech stimuli with longer tonic vowels (in Experiment 2) should be likewise difficult to classify. Similarly, if the advantage of non-speech stimuli in steps 1–5 (relative to the unspliced sixth step) results from their splicing, then these spliced steps should show a similar advantage in the present experiment. In contrast, if the speechiness of disyllables is due to their phonological well-formedness, then responses to non-speech stimuli in Experiment 2 should be unaffected by vowel duration. Such well-formedness effects, moreover, should also be evident in the overall pattern of responses to non-speech stimuli. Because the non-speech stimuli used in this experiment are all well-formed disyllables, we expect their structure to inhibit the non-speech response. Consequently, participants in Experiment 2 should experience greater difficulty in distinguishing non-speech stimuli from their speech-like counterparts.

Method

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

Materials and Procedure. The materials corresponded to the same three pairs of naturally produced disyllables used in Experiment 1. For each such disyllable, we gradually decreased the duration of the tonic vowel using a procedure identical to the one applied to the splicing of the pretonic vowel in Experiment 1. We first determined the portion of the tonic vowel slated for removal by a identifying a segment of 70 ms (12–14 pitch periods), matched to the duration of the pretonic vowel ($M = 68$ ms, ranging from 12 to 15 pitch periods). This segment was measured from the center of the vowel outwards, and it included some coarticulatory cues. The entire tonic vowel (unspliced) was presented in step 6. We next proceed to excise this segment in steady increments, such that steps 5–1 had 8, 6, 4, 2, and 0 pitch periods remaining out of the pitch periods slated for removal. Despite removing a chunk of the tonic vowel, the items in step 1 were clearly identified as disyllabic, and their remaining tonic vowel averaged 38 ms (7.16 pitch periods).

The resulting speech continua were next used to form non-speech and speech –control stimuli along the same method used in Experiment 1. The procedure was the same as in Experiment 1.

Results and Discussion

Outliers (responses faster than 2.5 SD from the means, less than 3% of all correct observations) were excluded from the analyses of response time. Mean response time for speech and non-speech stimuli as a function of the duration of the tonic vowel is presented in **Figure 3** (the accuracy means are provided in **Table 2**).

An inspection of means showed no evidence that responses to non-speech stimuli were monotonically linked to the duration of the tonic vowel. A 2 speech status × 2 continuum type × 6 vowel ANOVA yielded a reliable effect of continuum type [response accuracy: $F(1, 9) = 5.23$, MSE = 0.007, $p < 0.05$; response time: $F(1, 8) = 10.69$, MSE = 1413, $p < 0.02$], indicating that *md*-type stimuli were identified more slowly and less accurately than their *ml*-type counterparts. Because this effect did not depend on vowel duration, the difficulty with *md*-type stimuli is most likely due to the acoustic properties of those stimuli, rather than their phonological structure. The only other effect to approach significance was that of speech status (speech vs. non-speech) on response time [$F(1, 8) = 4.97$, MSE = 4757, $p < 0.06$]. No other effects were significant ($p > 0.17$).

Unlike Experiment 1, where speech stimuli were identified more readily than non-speech, in the present experiment, speech stimuli produced slower responses than their non-speech counterparts. This finding is consistent with the possibility that well-formed non-speech stimuli tend to be identified as speech, and consequently, they are harder to discriminate from non-speech-like inputs. Indeed the discrimination (d') of speech from non-speech was lower in Experiment 2 ($d' = 2.69$) relative to Experiment 1 ($d' = 3.82$).

Crucially, however, unlike Experiment 1, response to non-speech stimuli in the present experiment was not modulated by vowel duration. A separate 2 continuum type × 6 vowel duration

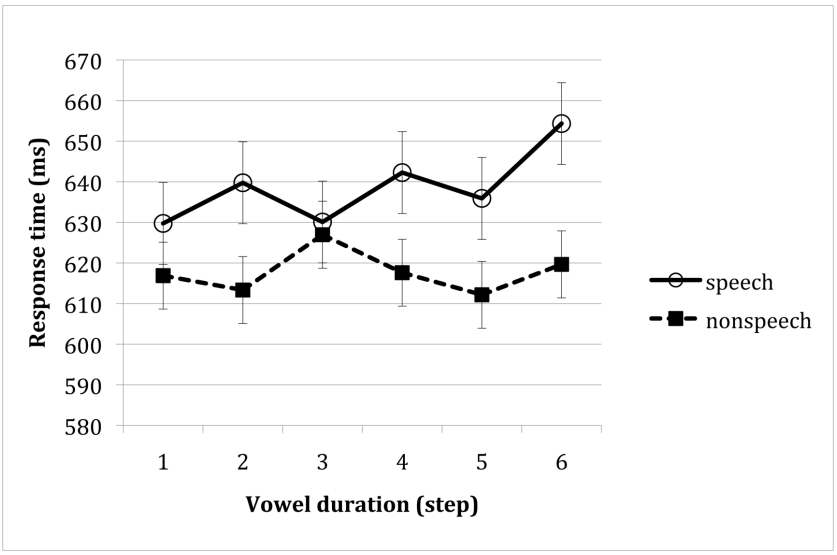


FIGURE 3 | Response time to speech and non-speech stimuli as a function of vowel duration (in Experiment 2). Error bars reflect confidence intervals, constructed for the difference among means along each of the vowel duration continua (i.e., speech and non-speech).

Table 2 | Mean response accuracy and response time to speech and non-speech stimuli in Experiment 2.

	Vowel duration	Stimulus type			
		Speech		Non-speech	
		ml	*md*	*ml*	*md*
Response accuracy	1	0.90	0.92	0.98	0.84
	2	0.91	0.90	0.91	0.84
	3	0.94	0.90	0.91	0.88
	4	0.96	0.92	0.9	0.9
	5	0.96	0.94	0.93	0.87
	6	0.91	0.90	0.92	0.87
Response time (ms)	1	631	628	610	624
	2	630	649	603	624
	3	603	657	621	633
	4	633	651	628	607
	5	630	641	602	622
	6	633	676	612	627
Mean (accuracy)		0.93	0.91	0.93	0.87
Mean (RT)		627	651	613	623

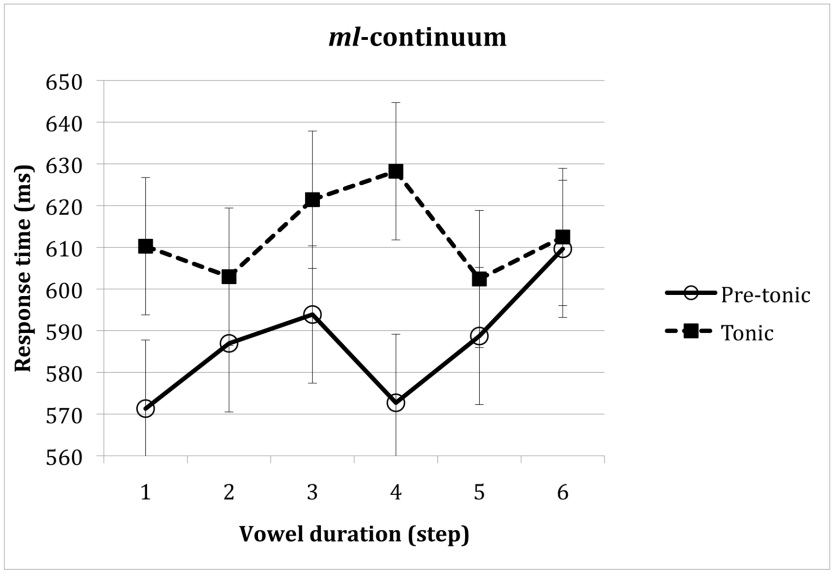
of the non-speech stimuli confirmed that responses to non-speech inputs were unaffected by vowel duration ($F < 1$, in response time and accuracy); the interaction also did not approach significance ($F < 1$, in response time and accuracy).

Given that the tonic vowel manipulation (in Experiment 2) closely matched the pretonic vowel manipulation (in Experiment 1), the confinement of the vowel effect to non-speech stimuli in Experiment 1 suggests that this effect specifically concerns the well-formedness of non-speech inputs, rather than vowel duration *per se*.

To further bolster this conclusion, we next compared the responses to non-speech across the two experiments using a 2 Experiment × 2 continuum type × 6 vowel duration ANOVA (see **Figure 4**)[1]. The analysis of response time yielded a significant effect of continuum type [$F(1, 17) = 6.08$, MSE = 977, $p < 0.03$] and a reliable three way interaction [$F(5, 85) = 2.42$, MSE = 1142, $p < 0.05$]. No other effects were significant (all $p > 0.19$). To interpret this interaction, we next examined responses to the two continuum types (*ml* vs. *md*) separately, using a 2 Experiment × 6 vowel duration ANOVAs. The analysis of the *ml*-continuum yielded no reliable effects (all $p > 0.14$). In contrast, the *md*-continuum yielded a marginally significant interaction [$F(5, 85) = 2.23$, MSE = 1392, $p < 0.06$]. No other effect was significant (all $p > 0.32$). We further interpreted the effect of vowel duration by testing for the simple main effect of vowel duration for the tonic vs. pretonic vowels, separately (in Experiment 2 vs. 1). Vowel duration was significant only for the pretonic vowel condition [$F(5, 45) = 5.47$, MSE = 746, $p < 0.0006$], but not in the tonic vowel condition [$F(5, 40) < 1$, MSE = 2118].

The attenuation of the tonic–pretonic contrast for the *ml*-continuum is likely due to phonetic factors. As noted earlier, the *md*-items exhibit a phonetic bifurcation due to the silence associated with the stop, and for this reason, disyllabicity might be more salient for *md*-items. The absence of a vowel effect in the *ml*-continuum indicates that merely increasing the duration of the vowel – whether it is tonic pretonic – is insufficient to impair the identification of non-speech stimuli. Results with the *md*-continuum, however, clearly show that vowel duration had

[1]Similar analyses conducted on the speech stimuli yielded only a significant effect of vowel duration [$F(5, 90) = 2.48$, MSE = 1361, $p < 0.04$] and a marginally significant effect of experiment [$F(1, 18) = 3.81$, MSE = 28538, $p < 0.07$] – no other effects were significant ($p > 0.13$). A comparison of the speech and non-speech stimuli across the two experiments (2 Experiment × 2 speech status × 2 continuum type × 6 vowel duration) yielded a reliable four-way interaction [$F(5, 85) = 2.84$, MSE = 1089, $p < 0.03$].



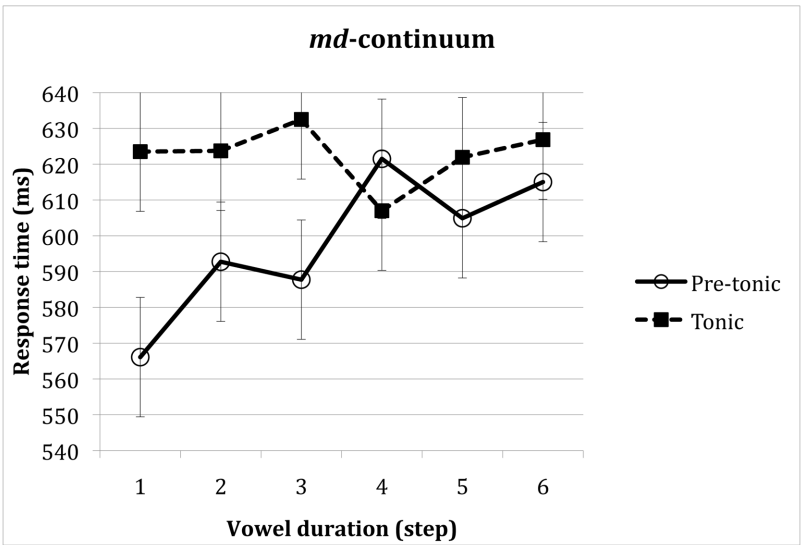


FIGURE 4 | The effect of the vowel manipulation (tonic vs. pretonic) on the identification of non-speech stimuli derived from the *ml*- vs. *md*-continua. Error bars reflect confidence intervals constructed for the difference between the means.

distinct effects on tonic and pretonic stimuli. While increasing the duration of the pretonic vowel impaired the identification of non-speech, the same increase in vowel length had no measurable effect when it concerned the tonic vowel – a phonetic contrast that does not affect well-formedness. The finding that identical vowel manipulations affected the identification of non-speech stimuli in a selective manner – only when it concerned the pretonic vowel, and only with the *md*-continuum – confirms that this effect is inexplicable by vowel duration *per se*. Merely increasing the duration of a vowel is insufficient to impair the classification of non-speech stimuli as such. Together, the findings of Experiments 1–2 suggest that well-formed structures are perceived as speech-like.

Experiments 3–4 further test this hypothesis by seeking converging evidence from an unrelated phenomenon in another language – Hebrew. To demonstrate that the effect of well-formedness on non-speech is not specific to conditions that require its comparison to edited speech stimuli (resynthesized or spliced), we compared non-speech stimuli with naturally produced speech. As in the case of English, we compared the ease of speech/non-speech discrimination for stimuli that were either phonologically well-formed or ill-formed. In the case of Hebrew, well-formedness is defined by the location of identical consonants in the stem – either initially (e.g., *titug*), where identical consonants are ill-formed in Semitic languages, or finally (e.g., *gitut*), where they are well-formed. Accordingly, the phonetic characteristics of

well-formedness differ markedly from the ones considered for English. To the extent that stimuli corresponding to well-formed structures are consistently harder to classify as non-speech (across different phonetic manifestations and languages), such a convergence would strongly implicate phonological structure as the source of this phenomenon.

PART 2: IDENTITY RESTRICTIONS ON HEBREW STEMS EXTEND TO NON-SPEECH STIMULI

Like many Semitic languages, Hebrew restricts the location of identical consonants in the stem: AAB stems (e.g., *titug*), where identical consonants occur at the left edge, are ill-formed, whereas their ABB counterparts (e.g., with identical consonants at the right edge, *gitut*) are well-formed (Greenberg, 1950). A large body of literature shows that Hebrew speakers are highly sensitive to this restriction and they freely generalize it to novel forms. Specifically, novel-AAB forms are rated as less acceptable than ABB counterparts (Berent and Shimron, 1997; Berent et al., 2001a), and because novel-AAB stems (e.g., *titug*) are ill-formed, people classify them as non-words more rapidly than ABB/ABC controls (e.g., *gitut, migus*) in the lexical decision task (Berent et al., 2001b, 2002, 2007b) and they ignore them more readily in Stroop-like conditions (Berent et al., 2005). Given that AAB Hebrew stems are clearly ill-formed, we can now turn to examine whether their structure might affect the classification of non-speech stimuli. If phonologically ill-formed stimuli are, in fact, more readily identifiable as non-speech, then, ill-formed AAB Hebrew stems should be classified as non-speech more easily than their well-formed (ABB and ABC) counterparts.

To examine this prediction, Experiment 3 compares the classification of three types of novel stems. Members of all three stem types are unattested in Hebrew, but they differ on their well-formedness. One group of stimuli, with an AAB (e.g., *titug*) structure is ill-formed, whereas the two controls – ABB (e.g., *gitut*) and ABC (e.g., *migus*) are well-formed. These items were recorded by a native Hebrew talker, and they were presented to participants in two formats: either unedited, as natural speech, or edited, such that they were identified as non-speech. Participants were asked to rapidly classify the stimulus as either speech or non-speech. If ill-formed stimuli are less speech-like, then non-speech stimuli with an AAB structure should elicit faster responses compared to their well-formed counterparts, ABB or ABC stimuli.

EXPERIMENT 3

Method

Participants. Twenty-four native Hebrew speakers, students at the University of Haifa, Israel, took part in the experiment for payment.

Materials. The materials corresponded to 30 triplets of speech stimuli along with 30 triplets of non-speech counterparts. All materials were non-words, generated by inserting novel consonantal roots (e.g., *ttg*) in the vocalic nominal template $C_1iC_2uC_3$ – the template of mishkal Piʔul (e.g., *ttg* + $C_1iC_2uC_3$ → *titug*). In each such triplet, one stem had identical consonants at its left edge (AAB), another had identical consonants at the right edge (ABB), and a third member (ABC) had no identical consonants (e.g., *titug*, *gitut*, *migus*). Within a triplet, AAB and ABB forms were matched for their identical consonants (e.g., *titug*, *gitut*) and the ABB and ABC forms were further matched for the co-occurrence of their consonants in Hebrew roots. The speech stimuli were recorded naturally, by a native Hebrew speaker – these materials were previously used in Berent et al. (2007b; Experiment 6), and they are described there in detail (see Berent et al., 2007b, Appendix A, for the list of stimuli). As noted there, the three types of stimuli did not differ reliably on their acoustic durations [$F < 1$; for AAB items: $M = 1191$ ms (SD = 108 ms); for ABB items: $M = 1195$ ms (SD = 103 ms); for ABC items: $M = 1171$ ms (SD = 101 ms)].

We next generated non-speech stimuli by adding together three synthetic sound components derived from the original stimulus waveforms. The first, low-frequency component was produced by lowpass filtering the stimulus waveforms at 400 Hz (slope of −85 dB per octave) to isolate the first formant, and deriving a spectral contour of the first formant frequency values from spectrograms of the filtered speech stimuli (256 point DFT, 0.5 ms time increment, Hanning window) using a peak-picking algorithm, which also extracted the corresponding amplitude values to produce an amplitude contour. Next, this low-frequency spectral contour was shifted up in frequency by multiplying it by 1.47, and then resynthesized into a sound component using a voltage-controlled oscillator modulated by the amplitude contour. The second, intermediate-frequency sound component was produced by bandpass filtering the original stimulus waveforms between 2000 and 4000 Hz (slope of −85 dB per octave), and deriving a single spectral contour of the frequency values in this intermediate range from spectrograms of the filtered speech stimuli (256 point DFT, 0.5 ms time increment, Hanning window) using a peak-picking algorithm, which also extracted the corresponding amplitude values to produce an amplitude contour. Next, this intermediate spectral contour was shifted down in frequency by multiplying it by 0.79, and then resynthesized into a sound component using a voltage-controlled oscillator modulated by the amplitude contour. The third, high-frequency sound component was produced by bandpass filtering the original stimulus waveforms between 4000 and 6000 Hz (slope of −85 dB per octave), and deriving a single spectral contour of the frequency values in this high range from spectrograms of the filtered speech stimuli (256 point DFT, 0.5 ms time increment, Hanning window) using a peak-picking algorithm, which also extracted the corresponding amplitude values to produce an amplitude contour. These three components were then summed together with relative amplitude ratios of 1.0:0.05:2.0 (low-frequency component: intermediate-frequency component: high-frequency component) to produce the non-speech version of each stimulus. The structure of these non-speech stimuli and their natural speech counterparts is illustrated in **Figure 5** (a sample of the materials is available at http://www.psych.neu.edu/faculty/i.berent/publications.htm). The experimental procedure was the same as in the previous experiments.

Results and Discussion

Figure 6 plots mean response time for speech and non-speech stimuli as a function of stem structure (the corresponding accuracy means are provided in **Table 3**). An inspection of the means suggests that speech and non-speech stimuli were readily identified

as intended. Moreover, well-formedness selectively modulated responses to non-speech stimuli: ill-formed AAB structures were identified more readily than their well-formed controls ABB and ABC given non-speech stimuli, but no such effect emerged with speech inputs.

These conclusions are borne out by the outcomes of the 2 speech status (speech/non-speech) × 3 stem-type (AAB/ABB/ABC) ANOVA. Since in this experiment, the conditions of interest are each represented by 30 different items, these analyses were conducted using both participants and items as random variables. The ANOVAs yielded a reliable interaction [In response time: $F(2, 46) = 6.30$, MSE = 1967, $p < 0.004$; $F(2, 58) = 7.78$, MSE = 2452, $p < 0.002$; In response accuracy: both $F < 1$] as well as a marginally significant effect of speech status [In response accuracy: $F_1(1, 23) = 3.03$, MSE = 0.036, $p < 0.10$, $F_2(1, 29) = 82.90$, MSE = 0.001, $p < 0.0001$; In response time: $F_1 < 1$; $F_2(1, 29) = 21.25$, MSE = 3235, $p < 0.00008$]. No other effect was significant (all $p > 0.14$). We next proceeded to interpret the interaction by testing for the simple main effects of stem structure and of speechiness, followed by planned orthogonal contrasts (Kirk, 1982).

Consider first the effect of stem structure. An analysis of non-speech stimuli yielded a reliable effect of stem structure in the analyses of response time [$F_1(2, 46) = 7.14$, MSE = 2312, $p < 0.003$; $F_2(2, 58) = 7.36$, MSE = 3257, $p < 0.002$]. Planned comparisons showed that responses to ill-formed, AAB stems were faster than their well-formed counterparts, either ABB [$t_1(46) = 2.31$, $p < 0.03$; $t_2(58) = 2.12$, $p < 0.04$] or ABC [$t_1(46) = 3.74$, $p < 0.0006$; $t_2(58) = 3.83$, $p < 0.0004$] stems, which, in turn, did not differ [$t_1(46) = 1.43$, $p > 0.15$, n.s.; $t_2(58) = 1.71$, $p > 0.09$]. Similar analyses conducted on speech materials produced no reliable effect of well-formedness (all $F < 1$). Thus, the advantage of ill-formed AAB stimuli was specific to non-speech inputs.

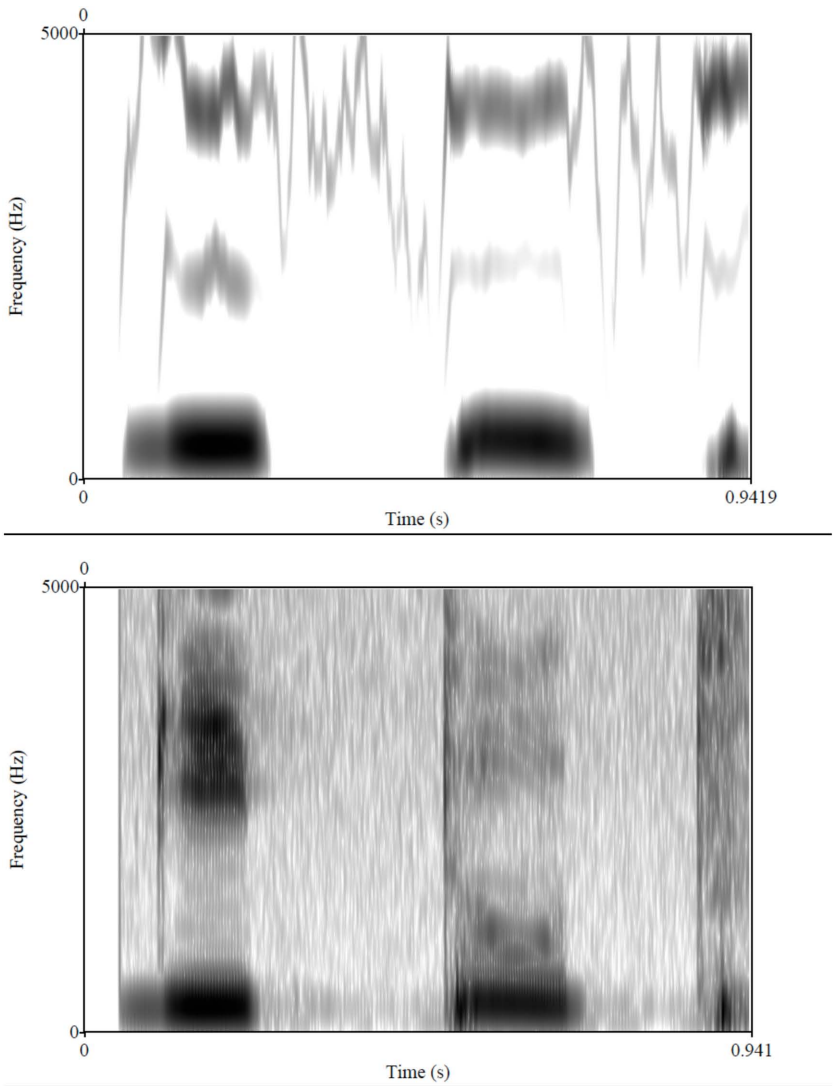


FIGURE 5 | Spectrograms for the novel Hebrew word "gitut." The top panel is a non-speech stimulus; the bottom one depicts natural speech.

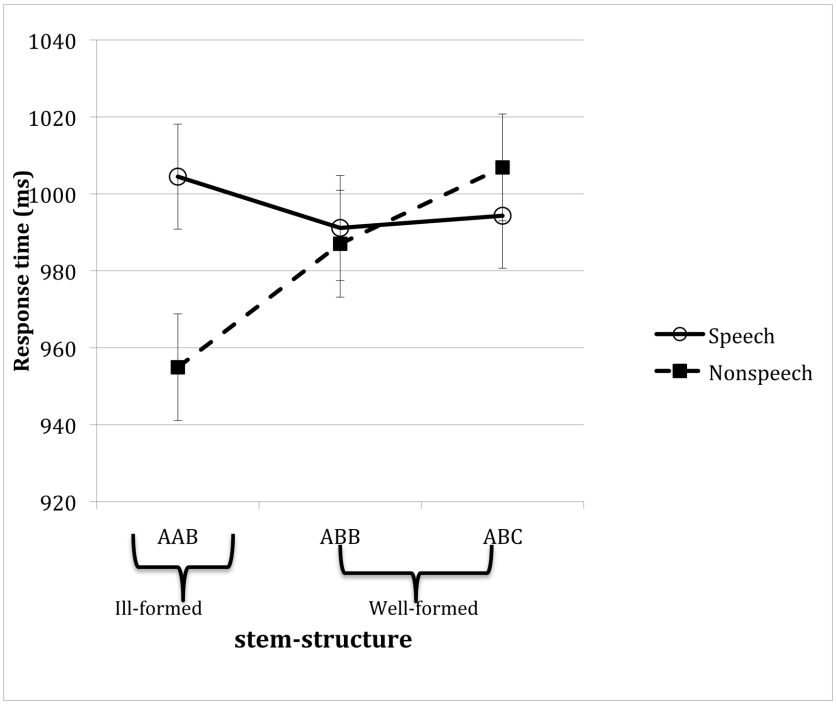


FIGURE 6 | Mean response time of Hebrew speakers to speech and non-speech inputs as a function of their phonological well-formedness in Hebrew. Error bars reflect confidence intervals constructed for the difference between the three types of stem structures, constructed separately for speech and non-speech stimuli.

Table 3 | Mean response accuracy (proportion correct) in Experiment 3.

Structure	Stimulus type	
	Speech	Non-speech
AAB	0.99	0.93
ABB	0.99	0.93
ABC	0.99	0.94

Tests of the simple main effect of speech status further indicated that non-speech stimuli promoted faster responses than speech stimuli given ill-formed AAB structures [$F_1(1, 23) = 5.56$, MSE = 5296, $p < 0.03$; $F_2(1, 29) = 31.03$, MSE = 2966, $p < 0.0001$], but not reliably so with their well-formed counterparts, either ABB [$F_1(1, 23) < 1$; $F_2(1, 29) = 4.10$, MSE = 3350, $p < 0.06$], or ABC (both $F < 1$) stems. These results demonstrate that ill-formedness facilitated the classification of non-speech stimuli.

EXPERIMENT 4

The persistent advantage of non-speech stimuli that are phonologically ill-formed across different structural manifestations and languages is clearly in line with our hypothesis that "speechiness" depends, *inter alia*, on phonological well-formedness. The fact that similar acoustic manipulations failed to produce the effect given well-formed stimuli (in Experiment 2) offers further evidence that the advantage concerns phonological structure, rather than acoustic attributes. Experiment 4 seeks to further dissociate phonological structure from the acoustic properties of ill-formed stimuli using a complementary approach. Here, we maintained the acoustic properties by using the same stimuli as Experiment 3, but we altered their phonological well-formedness by presenting these items to a group of English speakers. English does not systematically restrict the location of identical consonants in stems, and our past research suggested that, to the extent English speakers constrain the location of identical consonants, their preference is opposite to Hebrew speakers', showing a slight preference for AAB forms (Berent et al., 2002, footnote 7). Clearly, English speakers should not consider AAB items ill-formed. If the tendency of Hebrew speakers to classify AAB stems as non-speech-like is due to the acoustic properties of these items, then the results of English should mirror the Hebrew participants. If, in contrast, the easier classification of AAB stimuli as speech is due to their phonological structure, then the findings from English speakers should diverge with Hebrew participants.

Method

Participants. Twenty-four English speakers, students at Northeastern University, took part in this study in partial fulfillment of a course requirement.

The materials and procedure were identical to Experiment 3.

Results

Mean response time and response accuracy to speech and non-speech stimuli is provided in **Figure 7** (the accuracy means are listed in **Table 4**). An inspection of the means suggests that, unlike Hebrew speakers, English participants' responses to non-speech

stimuli were utterly unaffected by stem structure. Stem structure, however, did modulate responses to speech stimuli, such that stems with identical consonants produced faster responses than no-identity controls.

A 2 speech status (speech/non-speech) × 3 stem-type (AAB/ABB/ABC) ANOVA indeed yielded a marginally significant interaction in the analyses of response time [$F_1(2, 46) = 3.14$, MSE = 772, $p < 0.06$; $F_2(2, 58) = 1.08$, MSE = 2409, $p < 0.34$; in response accuracy, both $F < 1$). The same ANOVA did not yield a reliable effect of stem-type [In response time: $F_1(2, 46) = 3.67$, MSE = 976, $p < 0.04$; $F_2(2, 58) = 1.88$, MSE = 2609, $p < 0.17$; In accuracy: both $F < 1$], but speech status was found to reliably modulate response accuracy [In response time: $F_1(1, 23) = 3.08$, MSE = 9710, $p < 0.10$; $F_2(1, 29) = 15.42$, MSE = 3745, $p < 0.0003$; In accuracy: $F_1(1, 23) = 5.83$, MSE = 0.03, $p < 0.03$; $F_2(1, 29) = 33.20$, MSE = 0.0009, $p < 0.00004$].

A separate analysis of the response latency to non-speech stimuli confirmed that the ability of English speakers to classify non-speech stimuli was utterly unaffected by stem structure (both $F < 1$). Stem structure, however, did modulate response to speech inputs [$F_1(2, 46) = 6.32$, MSE = 940, $p < 0.004$; $F_2(2, 58) = 3.28$, MSE = 2391, $p < 0.05$). Planned contrasts further suggested that responses to speech-ABC stems were significantly slower than ABB inputs [$t_1(46) = 3.44$, $p < 0.002$; $t_2(58) = 2.52$, $p < 0.02$], and marginally so relative to AAB ones [$t_1(46) = 2.49$, $p < 0.02$; $t_2(58) = 1.68$, $p < 0.10$], which, in turn, did not differ (both $t < 1$).

Discussion

The findings from Experiment 4 demonstrate that the processing of non-speech stimuli is modulated by linguistic knowledge. While Hebrew participants (in Experiment 3) responded reliably faster to non-speech AAB stems – stems that are ill-formed in their language, English participants in the present experiment were utterly insensitive to the structure of the same non-speech stimuli. And indeed, English does not systematically constrain the location of identical consonants in the stem. The selective sensitivity of Hebrew, but not English speakers to the structure of non-speech stimuli demonstrates that this effect reflects linguistic knowledge, rather than the acoustic properties of those stimuli.

While stem structure did not affect the responses of English participants to non-speech inputs, it did modulate their responses to speech stimuli: speech stimuli with identical consonants – either AAB or ABB – were identified faster than ABC controls. Since English speakers did not differentiate AAB and ABB stems, this effect must be due to reduplication *per se*, rather than to its location. Indeed, many phonological systems produce identical consonants by a productive grammatical operation of reduplication (McCarthy, 1986; Yip, 1988; Suzuki, 1998). It is thus conceivable that English speakers encode AAB and ABB stems as phonologically structured, and consequently, they consider stems as better formed than no-identity controls. The sensitivity of English speak-

Table 4 | Mean response time and response accuracy in Experiment 4.

Structure	Response accuracy (proportion correct)	
	Speech	Non-speech
AAB	0.99	0.96
ABB	0.99	0.95
ABC	0.99	0.97

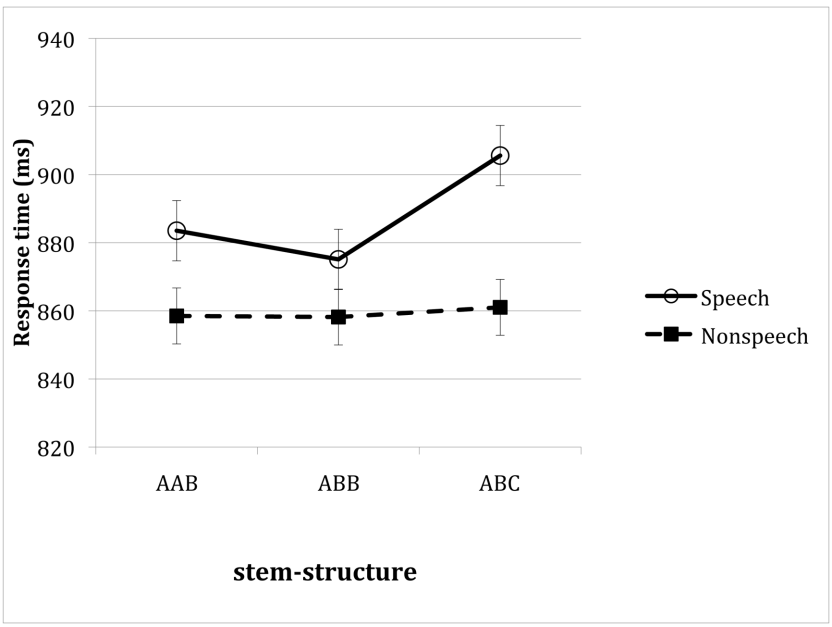


FIGURE 7 | Mean response time of English speakers to speech and non-speech inputs as a function of their stem structure. Error bars reflect confidence intervals constructed for the difference between the three types of stem structures, constructed separately for speech and non-speech stimuli.

ers to consonant-reduplication is remarkable for two reasons. First, reduplication is not systematically used in English, so the sensitivity of English speakers to reduplication might reflect the encoding of a well-formedness constraint that is not directly evident in their own language (for other evidence consistent with this possibility, see Berent et al., 2007a). Second, the finding that reduplicated (AAB and ABB) stems are more readily recognized as speech suggests that well-formedness affects not only non-speech stimuli but also the processing of speech inputs.

GENERAL DISCUSSION

Much research suggests that people can extract linguistic messages from auditory carriers that they classify as "non-linguistic" (e.g., Remez et al., 1981, 2001). Here, we examine whether structural aspects of linguistic messages can inform the classification of these messengers as linguistic. Four experiments gauged the effect of phonological well-formedness on the discrimination of non-speech stimuli from various speech controls. In Experiment 1, we showed that English speakers experience difficulty in the classification of non-speech stimuli generated from syllables that are phonologically well-formed (e.g., *melif*) compared to ill-formed counterparts (e.g., *mlif*). Experiment 3 replicated this effect using a second manifestation of well-formedness in a different language – the restrictions on identical consonants in Hebrew stems. Once again, participants (Hebrew speakers) experienced difficulties responding to non-speech stimuli that are well-formed in their language (e.g., *gitut*) compared to ill-formed controls (e.g., *titug*). The converging difficulties across diverse manifestations of well-formedness suggest that these effects are likely due to phonological structure, rather than the acoustic properties of these stimuli.

Experiments 2 and 4 further support this conclusion by showing that acoustic manipulations similar to the ones in Experiments 1 and 3, respectively, fail to produce such difficulties once well-formedness is held constant. Specifically, Experiment 2 showed that merely increasing the duration of a vowel (a manipulation that mimics the *mlif–məlif* contrast from Experiment 1) is insufficient to impair the classification of non-speech stimuli once long- and short-vowel items are both well-formed in participants' language (English). Similarly, Experiment 4 showed that non-speech items that are well-formed in Hebrew present no difficulties for English participants. The convergence across two different manipulations of well-formedness, on the one hand, and its divergence with the outcomes of similar (or even identical) acoustic manipulations, on the other, suggest that the observed difficulties with well-formed non-speech stimuli are due to productive linguistic knowledge. As such, these results suggest that structural properties of linguistic messages inform the classification of acoustic messengers as linguistic.

While our present results do not directly speak to the nature of the knowledge consulted by participants, previous findings suggest that it is inexplicable by familiarity – either statistical knowledge or familiarity with the coarse acoustic properties of the language (e.g., familiarity accounts such as Rumelhart and McClelland, 1986; Goldinger and Azuma, 2003; Iverson et al., 2003; Iverson and Patel, 2008; Yoshida et al., 2010). Recall, for example, that well-formed nasal-initial sequences exhibited a stronger speechiness effect even when the stimuli were unfamiliar, as these particular items (*mdif*-type monosyllables) – while structurally well-formed – happened to be unattested in the participants' language (Russian; Berent et al., 2010). Likewise, the restriction on identical Hebrew consonants generalizes across the board, to novel segments and phonemes (e.g., Berent et al., 2002), and computational simulations have shown that such generalizations fall beyond the scope of several non-algebraic mechanisms (Marcus, 2001; Berent et al., in press). The algebraic properties of phonological generalizations, on the one hand, and their demonstrable dissociation with acoustic familiarity, on the other, suggest that the knowledge available to participants specifically concerns grammatical well-formedness[2]. Whether such algebraic knowledge modulates the identification of non-speech stimuli, specifically, remains to be seen. But regardless of what linguistic knowledge is consulted in this case, it is clear that some structural attributes of the linguistic message inform the classification of auditory stimuli as speech.

Why are well-formed non-speech stimuli harder to classify? Earlier, we proposed two possible loci for the effect of phonological well-formedness. One possibility is that well-formedness affects the allocation of attention to acoustic stimuli. In this account, well-formed stimuli engage attentional resources that are necessary for the speech-discrimination task, and consequently, the classification of non-speech stimuli suffers compared to ill-formed structures. On a stronger interactive account, well-formedness informs the evaluation of acoustic inputs by the language system itself. In this view, the output of the phonological grammar feeds back into the evaluation of the input, such that better-formed inputs are interpreted as more speech-like. While the weak attention view and strong interactive accounts both predict difficulties with well-formed non-speech stimuli, they differ with respect to their predictions for speech inputs. The strong interactive account predicts that well-formed speech stimuli should be easier to recognize as speech, whereas the attention-grabbing explanation predicts that well-formed speech stimuli should likewise engage attention resources, hence, they should be harder to classify than ill-formed counterparts.

While our results are not entirely conclusive on this question, two observations favor the stronger interactive perspective. First, well-formedness impaired the identification of non-speech stimuli in Experiment 1, but it had no such effect on speech stimuli. The relevant (speech status × vowel duration) interaction, however, was not significant, so the interpretation of this finding requires some caution. Stronger differential effects of well-formedness obtained in Experiment 3. Here, well-formedness selectively impaired the classification of non-speech stimuli. Moreover, well-formedness produced the opposite effect on speech inputs (in Experiment 4): well-formed inputs with reduplication were identified more readily as speech. Although these conclusions are limited in as much as the contrasting findings for speech and non-speech stimuli come from different experiments (Experiments 3 vs. 4), these

[2]Note that our conclusions concern cognitive architecture, not its neural instantiation. The account outlined here is perfectly consistent with the possibility that phonological knowledge reshapes auditory brain areas, including low-level substrates. At the functional level, however, such changes support generalizations that are discrete and algebraic.

results are nonetheless more consistent with the proposal that well-formedness facilitates the classification of acoustic stimuli as speech. Accordingly, the status of a stimulus as "linguistic" appears to depend not only on its inherent acoustic attributes, but also on the structure of the outputs it affords.

The finding that better-formed linguistic structures are more readily classified as speech suggests that the outputs of the language system (i.e., structural description) can penetrate the processing of its inputs (i.e., the classification of acoustic stimuli as speech). Top-down effects on speech classification are not new (e.g., Remez et al., 1981). Our results, however, are the first to demonstrate that these effects can originate from structural descriptions computed by the language system itself. Although these findings leave open the possibility that the language system is specialized with respect to structures that it can compute, they do suggest that the selection of its inputs is not encapsulated.

ACKNOWLEDGMENT

This research was supported by NIDCD grant DC003277 to Iris Berent. We wish to thank Tracy Lennertz and Katherine Harder for their help with the preparation of the auditory stimuli, and Monica Bennett, for technical assistance.

REFERENCES

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., and Menon, V. (2010). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* 21, 1507–1518.

Azadpour, M., and Balaban, E. (2008). Phonological representations are unconsciously used when processing complex, non-speech signals. *PLoS ONE* 3, e1966. doi: 10.1371/journal.pone.0001966

Berent, I., Balaban, E., Lennertz, T., and Vaknin-Nusbaum, V. (2010). Phonological universals constrain the processing of nonspeech. *J. Exp. Psychol. Gen.* 139, 418–435.

Berent, I., Bibi, U., and Tzelgov, J. (2005). The autonomous computation of linguistic structure in reading: evidence from the Stroop task. *Mental Lex.* 1, 201–230.

Berent, I., Everett, D. L., and Shimron, J. (2001a). Do phonological representations specify variables? Evidence from the obligatory contour principle. *Cogn. Psychol.* 42, 1–60.

Berent, I., Shimron, J., and Vaknin, V. (2001b). Phonological constraints on reading: evidence from the obligatory contour principle. *J. Mem. Lang.* 44, 644–665.

Berent, I., and Lennertz, T. (2010). Universal constraints on the sound structure of language: phonological or acoustic? *J. Exp. Psychol. Hum. Percept. Perform.* 36, 212–223.

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

Berent, I., Marcus, G. F., Shimron, J., and Gafos, A. I. (2002). The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition* 83, 113–139.

Berent, I., and Shimron, J. (1997). The representation of Hebrew words: evidence from the obligatory contour principle. *Cognition* 64, 39–72.

Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (2007a). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104, 591–630.

Berent, I., Vaknin, V., and Marcus, G. (2007b). Roots, stems, and the universality of lexical representations: evidence from Hebrew. *Cognition* 104, 254–286.

Berent, I., Wilson, C., Marcus, G., and Bemis, D. (in press). On the role of variables in phonology: remarks on Hayes and Wilson. *Linguist. Inq.* 43.

Brentari, D., Coppola, M., Mazzoni, L., and Goldin-Meadow, S. (2011). When does a system become phonological? Handshape production in gestures, signers and homesigners. *Nat. Lang. Linguist. Theory*. doi: 10.1007/s11049-011-9145-1. [Epub ahead of print].

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Dehaene-Lambertz, G., Pallier, C., Semiclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Goldinger, S. D., and Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *J. Phon.* 31, 305–320.

Greenberg, J. H. (1950). The patterning of morphemes in semitic. *Word* 6, 162–181.

Iverson, J. R., and Patel, A. D. (2008). Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* 124, 2263–2271.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57.

Kirk, R. (1982). *Experimental Design: Procedures For the Behavioral Sciences*. Pacific grove: Brooks/Cole.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461.

Liberman, A. M., and Mattingly, I. G. (1989). A specialization for speech perception. *Science* 243, 489–494.

Liebenthal, E., Binder, J. R., Piorkowski, R. L., and Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *J. Cogn. Neurosci.* 15, 549–558.

Lukatela, G., Eaton, T., Sabadini, L., and Turvey, M. T. (2004). Vowel duration affects visual word identification: evidence that the mediating phonology is phonetically informed. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 151–162.

Lukatela, G., Eaton, T., and Turvey, M. T. (2001). Does visual word identification involve a sub-phonemic level? *Cognition* 78, B41–B52.

Maddieson, I. (2006). "In search of universals," in *Linguistic Universals*, eds R. Mairal and J. Gil (Cambridge: Cambridge University Press), 80–100.

Marcus, G. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge: MIT press.

McCarthy, J. J. (1986). OCP effects: gemination and antigemination. *Linguist. Inq.* 17, 207–263.

Meyer, M., Zaehle, T., Gountouna, V.-E., Barron, A., Jancke, L., and Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport* 16, 1985–1989.

Molfese, D. L., and Molfese, V. J. (1980). Cortical response of preterm infants to phonetic and nonphonetic speech stimuli. *Dev. Psychol.* 16, 574–581.

Pinker, S. (1994). *The Language Instinct*. New York: Morrow.

Pinker, S., and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition* 95, 201–236.

Prince, A., and Smolensky, P. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell Pub.

Prince, A., and Smolensky, P. (1997). Optimality: from neural networks to universal grammar. *Science* 275, 1604–1610.

Remez, R. E., Pardo, J. S., Piorkowski, R. L., and Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychol. Sci.* 12, 24–29.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949.

Rogalsky, C., Rong, F., Saberi, K., and Hickock, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 2843–3852.

Rumelhart, D. E., and McClelland, J. L. (1986). "On learning past tense of english verbs: implicit rules or parallel distributed processing?" in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2, eds D. E. Rumelhart, J. L. McClelland, and T. P. R. Group (Cambridge, MA: MIT Press), 216–271.

Sandler, W., Aronoff, M., Meir, I., and Padden, C. (2011). The gradual emergence of phonological form in a new language. *Nat. Lang. Linguist. Theory* 29, 505–543.

Sandler, W., and Lillo-Martin, D. C. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.

Shultz, S., and Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. *Lang. Learn. Dev.* 6, 241–257.

Suzuki, K. (1998). *A Typological Investigation of Dissimilation*. Tucson, AZ: University of Arizona.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., and Wartenburger, I. (2009). Sensitivity to newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* 29, 14726–14733.

Trout, J. (2003). Biological specialization for speech: what can the animals tell us? *Curr. Dir. Psychol. Sci.* 12, 155–159.

Van Orden, G. C., Pennington, B. F., and Stone, G. O. (1990). Word

identification in reading and the promise of subsymbolic psycholinguistics. *Psychol. Rev.* 97, 488–522.

Vouloumanos, A., Hauser, M. D., Werker, J. F., and Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Dev.* 81, 517–527.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., and Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* 13, 994–1005.

Vouloumanos, A., and Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* 10, 159–164.

Yip, M. (1988). The obligatory contour principle and phonological rules: a loss of identity. *Linguist. Inq.* 19, 65–100.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., and Werker, J. F. (2010). The development of perceptual grouping biases in infancy: a Japanese-English cross-linguistic study. *Cognition* 115, 356–361.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 March 2011; paper pending published: 11 May 2011; accepted: 19 July 2011; published online: 13 September 2011. Citation: Berent I, Balaban E and Vaknin-Nusbaum V (2011) How linguistic chickens help spot spoken-eggs: phonological constraints on speech identification. Front. Psychology **2**:*182. doi: 10.3389/fpsyg.2011.00182*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

Copyright © 2011 Berent, Balaban and Vaknin-Nusbaum. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.