Published in final edited form as:

Atten Percept Psychophys. 2013 January ; 75(1): 101–120. doi:10.3758/s13414-012-0376-y.

Perceptual representations of phonotactically illegal syllables

Mara Breen¹, John Kingston², and Lisa D. Sanders¹

¹University of Massachusetts, Department of Psychology, Neuroscience and Behavior Program

²University of Massachusetts, Department of Linguistics

Abstract

Listeners often categorize phonotactically illegal syllables (e.g., /dla/ in English) as phonemically similar legal ones (e.g., /gla/). In an earlier investigation of this effect, Dehaene-Lambertz, Dupoux, and Gout (2000) did not observe a mismatch-negativity in response to deviant, illegal syllables, and therefore argued that phonotactics constrain early perceptual processing. In the current study, using a priming paradigm, we compared event-related potentials elicited by legal targets (/gla/) preceded by (1) different tokens of identity primes (/gla/), (2) phonemically distinct control primes (/kla/), and (3) phonotactically illegal primes (/dla/). Targets elicited a larger positivity 200–350 ms after onset when preceded by illegal primes compared to identity primes. Later portions of the waveforms (350–600 ms) did not differ for targets preceded by identity and illegal primes, and corresponded to similarity ratings. These data support a model of speech perception in which phoneme clusters are faithfully represented at early stages of perception and only later are categorized according to language-specific phonotactic constraints.

Keywords

speech perception; phonotactics; event-related potentials; priming

Representations of non-native speech sounds and patterns are strongly influenced by native language experience. In particular, listeners assimilate non-native sounds to native ones. As a result, patterns that do not exist in their language are reported to be ones that do (Brown & Hildum, 1956; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; Hallé & Best, 2007; Hallé, Best, & Bachrach, 2003; Hallé, Segui, Frauenfelder, & Meunier, 1998; Massaro & Cohen, 1983; Moreton, 2002). The goal of the current study was to determine *when* in the course of processing this assimilation takes place.

Perceptual Assimilation

Perceptual assimilation of non-native to native patterns has been reported for a variety of languages. One example comes from Japanese, which requires that adjacent consonants with different places of articulation be separated by a vowel. Illegal consonant clusters in loan

words from other languages are repaired by inserting a vowel between the consonants. For example, “MacDonald’s” becomes *makudonarudo* where the underlined vowels are inserted between /k/ and /d/, /r/ and /d/, and after the final /d/ to make the word phonotactically legal. This phonotactic constraint influences reported perception in addition to production. In response to an /ebuzo/ to /ebzo/ continuum constructed by incrementally removing the /u/ between the /b/ and /z/, the percentage of trials on which Japanese listeners reported that the vowel was present dropped from 100% to only 70%. In contrast, French listeners’ vowel-present responses across the continuum were veridical, dropping from 100% to 5% (Dupoux et al., 1999) – French differs from Japanese in permitting consonants like /b/ and /z/ to occur next to one another. These results suggest that the constraints on Japanese syllable structure cause native speakers to perceive the vowel even in its absence. Further evidence indicates that the source of the inserted vowel is Japanese phonotactics rather than its lexicon (Dupoux et al., 2001).

Another example of the influence of speakers’ native language phonotactics on their processing of non-native speech patterns comes from English. In English, the clusters /tʌ/ and /stʌ/ can never occur in the same syllable, while the clusters /tr/ and /sl/ are legal and frequent, as in *trip* and *slip*. Massaro and Cohen (1983) presented an /r-l/ continuum in the contexts /t_i/ and /s_i/ to English listeners, who identified the ambiguous tokens in the continuum more often as “r” after /t/ and more often as “l” after /s/. In both cases, the context resulted in the ambiguous sounds being reported more often as members of the phonotactically legal category.

The prohibition against /tʌ/ in English extends to /dʌ/ as well, and both clusters are also prohibited in French. When presented with the clusters /dʌ, tʌ/ at the beginning of a syllable, English and French listeners report /gl, kl/ and they fail to discriminate /dʌ, tʌ/ from /gl, kl/ (Hallé & Best, 2007; Hallé et al., 2003; Hallé et al., 1998; Massaro & Cohen, 1983; Moreton, 2002).¹ Not all languages ban /dʌ/ and /tʌ/ at the beginning of syllables. For example, they are legal in Hebrew, where *ṭalim* “dews” contrasts with *kalim* “rules”. Unlike English and French listeners, Hebrew listeners have no difficulty identifying /d, t/ in front of /l/ or in discriminating /d, t/ from /g, k/ in this context (Hallé & Best, 2007). The legality of these clusters in Hebrew is important because the stimuli employed in the current study were produced by native Hebrew speakers.

The non-native assimilation demonstrated by Japanese, English, and French speakers clearly shows that the phonotactics of listeners’ native languages influence their processing of strings of segments. However, this evidence does not indicate *when* in the course of processing the native language’s phonotactic influence occurs. Does phonotactic knowledge alter listeners’ early representations of the segment strings or only their later categorization of its contents? That is, are illegal consonant clusters separated by a vowel in Japanese listeners’ earliest measurable auditory representations, as Dupoux et al. (1999, 2001) and Dehaene-Lambertz et al. (2000) assert, and are [d, t] represented as /g, k/ before /l/ in English and French listeners’ earliest auditory representations? The *autonomous* hypothesis

¹The /d, t/ in English words like *bedlam*, *Hadley*, *antler*, and *Atlantic* are not misreported because they occur in different syllables than the /l/ (Moreton, 2002).

asserts that speech sounds are always perceived faithfully regardless of whether they occur in the language or are legal in the context in which they occur (Norris, McQueen, & Cutler, 2000). According to this hypothesis, phonotactic violations are repaired late in processing. In the *interactive* alternative (McClelland & Elman, 1986), the repairs to non-native speech patterns are instead immediate.

Event-related Potentials

The off-line behavioral measures reported in the studies cited above reveal little about the time course of the repair process, and therefore cannot adjudicate between the autonomous and interactive views. Event-related brain potentials (ERPs) make it possible to distinguish between primarily perceptual levels of representation (i.e., less than 350 ms after sound onset) and later higher-level representations (i.e., more than 350 ms after sound onset). For example, in a typical oddball paradigm, the repetition of a standard sound is interrupted infrequently and randomly by a deviant sound or sounds. Relative to the waveform elicited by the standard stimulus, the waveform elicited by the deviant exhibits a mismatch-negativity (MMN) that peaks between 150 and 350 ms after sound onset. The MMN is commonly assumed to reflect the automatic updating of auditory working memory when listeners are confronted with new, deviant input (Näätänen, Paavilainen, Rinne, & Alho, 2007).

Näätänen et al. (1997) presented Estonian and Finnish speakers with a series of speech sounds, in which the vowel /e/ was the standard and the vowels /o/, /ö/, or /õ/ were deviants. The first two deviants /o/ and /ö/ contrast with the standard /e/ in both languages, but /õ/ only contrasts with /e/ in Estonian. An MMN was observed in response to all three deviants in Estonian listeners, but only for the first two in Finnish listeners. This result was interpreted as evidence that the MMN is sensitive to native-language categories. A similar difference was reported for a Finnish vowel contrast between Hungarian speakers who were fluent in Finnish and showed an MMN versus other Hungarian speakers who had no prior exposure to the language and did not show an MMN (Winkler et al., 1999).

Dehaene-Lambertz et al. (2000) used a modified oddball design to investigate the perception of adjacent consonants by Japanese and French listeners. As stated earlier, Japanese requires that adjacent consonants with different places of articulation be separated by a vowel, while French has no such requirement; therefore, sequences like /gm/ are illegal in Japanese, but legal in French.² They presented listeners with sequences of five stimuli, consisting of four tokens of a /VCCV/ or /VCuCV/ string and either a fifth token of the same string or a token of the other string. For example, the deviant string at the end of the sequence was /igumo/ if the first four stimuli were /igmo/ and /igmo/ if the first four were /igumo/. In this trial structure, the fifth sound was a standard if it had the same sequence of segments as the previous stimuli and was a deviant if a vowel had been inserted between or removed from between the two consonants. French listeners distinguished the deviants from standards but Japanese listeners did not. In addition, the deviant stimulus elicited an MMN for the French listeners, but not for the Japanese listeners. Dehaene-Lambertz et al. interpreted this result as

²In this experiment, the adjacent consonants always crossed syllable boundaries.

evidence that a phonotactic filter repairs the non-native consonant clusters to conform to native language rules, and that this repair is effected at the earliest stages of processing in which the brain's responses were measured. These results, as well as those reported by Näätänen et al. (1997) and Winkler et al. (1999), thus support the interactive hypothesis.

The Current Study

The current study was also designed to investigate the levels of representation that are repaired for phonotactic violations. However, it differs methodologically in two important ways from Dehaene-Lambertz et al.'s study. First, their conclusion that listeners immediately repair illegal phonotactic patterns to legal ones rests on a between-subjects comparison. Specifically, Japanese listeners demonstrated no early ERP difference, but French listeners did. However, this result could be due to Japanese listeners explicitly reporting a deviant sound on far fewer trials than the French listeners – Japanese listeners reported that the fifth stimulus was different on only 8.6% of deviant trials compared to French listener's 95.1%. This difference in behavioral responses could induce a bias toward "same" responses and potentially reduce Japanese listeners' attention and sensitivity to stimulus differences. Therefore, in the current study, we tested the time course of perceptual assimilation by comparing the brain responses of the same individuals to patterns that are both legal and illegal.

A second difference between the current study and that of Dehaene-Lambertz et al. (2000) is that we did not adopt an oddball design. Some recent research suggests that the MMN may not be the best indicator of auditory memory (Schröger, 2007), and it has been shown that MMN amplitude can be modulated by attention (Sussman, 2007). In contrast, a large body of literature consistently shows that shared characteristics of prime-target pairs result in decreased ERP amplitude elicited by targets (e.g. Coch, Grossi, Coffey-Corina, Holcomb, & Neville, 2002; Coch, Grossi, Skendzel, & Neville, 2005; Grossi, Coch, Coffey-Corina, Holcomb, Neville, 2001; Weber-Fox, Spencer, Cuadrado, & Smith, 2003). These two considerations led us to use a priming design in the current study.

Listeners heard two stimuli on each trial – a prime followed by a target; they were instructed to rate the similarity of the two. The prime-target pairs used in the current experiment are listed in Table 1. In all combinations of Trial Type and Prime Status, the target was phonotactically legal in English. We compared ERPs elicited by these physically identical target syllables preceded by Identity, Control, or Test primes, with the following predictions. First, a target /gla/ preceded by a different token of the Identity prime /gla/ should result in the largest amount of priming and thus the smallest ERP amplitude; moreover, subjects should rate the prime and target as similar. Second, the Control prime /kla/, which differs from the target /gla/ by one contrastive feature (/k/ is voiceless and /g/ is voiced), should prime the target to a lesser extent. In addition, because the initial stop in the Control primes is phonotactically legal in the cluster in which it occurs, we predicted participants would rate the prime and target as different in this condition.

The critical condition is that of the Test prime, and the possible results can adjudicate between the interactive and autonomous hypotheses. The Illegal Test prime /dla/ also differs

from the target /gla/ by just one feature (/d/ is alveolar and /g/ is velar). If listeners represent /d/ as /g/ before /l/, as the behavioral evidence cited above and the interactive hypothesis would predict, then /dla/ should prime /gla/ just as well as the Identity prime /gla/, and participants should rate the prime and target as similar. If ERPs elicited by the targets following Illegal Test and Identity primes are similarly small, the interactive hypothesis is supported. If, on the other hand, participants represent the difference between /d/ and /g/ in this context, the Test prime should prime the target to a lesser extent than the Identity prime, and participants should rate the prime and target as different. If ERPs elicited by the targets following Illegal Test and Identity primes are distinct, the autonomous alternative is supported. The autonomous hypothesis further predicts that similarities in the Illegal Test and Control prime conditions will occur during perceptual processing less than 350 ms after target onset. In contrast, later portions of the ERP waveform should be similar for the Illegal Test and Identity prime conditions, reflecting participants' categorization of the clusters as combinations that are legal in their native language and corresponding to the explicit, behavioral similarity ratings of primes and targets.

A second goal of this study was to determine whether the representations of syllables prohibited by a phonotactic constraint (e.g. /dla/) are qualitatively different from those of syllables which are not illegal, but are (a) absent, because they have never arisen during the language's historical development, or are (b) rare but legal. In English, the former category is exemplified by clusters of a bilabial stop /b, p/ followed by /w/, which are absent except in loan words like *Puerto Rico*, *pueblo*, or *bueno*. The latter is exemplified by clusters of an alveolar stop /d, t/ followed by /w/, which occur relatively rarely in words like *dwarf* and *twice*. The stimuli used to address this second question were adapted from Moreton (2002), who demonstrated that, although /bw/ and /dl/ are both absent in English, listeners only have a phonotactic bias against perceiving /dl/. These stimuli are also presented in Table 1.

METHOD

Participants

Eighteen adult participants (4 female), ranging in age from 20–30 years ($M = 24.5$ years) provided data included in the analysis. All participants reported being right-handed, having normal hearing, normal or corrected-to-normal vision, taking no psychoactive medications, and having no neurological disorders. Participants were all monolingual native speakers of American English, with no experience with Hebrew, the language from which the stimuli were drawn. Five additional participants contributed data to the experiment; data from 2 were discarded due to recording error, from 2 because of low-frequency drift caused by skin potentials, and from 1 because of high frequency noise. All participants provided written consent and were compensated \$10/hr for their participation.

Stimuli

We manipulated 4 factors with 3, 3, 2, and 2 levels, respectively, resulting in 36 conditions. Examples of each condition are shown in Table 1. The first factor was Trial Type, defined by the shared characteristics of the prime and target. The pairs were: (a) different tokens of

phonologically identical syllables (e.g., prime = /gla/, target = /gla/), (b) syllables that differed in one feature of the first consonant and formed two legal sequences, (e.g., prime = /kla/, target = /gla/), or (c) syllables that differed by a single feature in the first consonant such that the prime met specific phonotactic conditions (e.g., prime = /dla/, target = /gla/). These three Trial Types are referred to as Identity, Control, and Test, respectively. The second factor was the linguistic status of the prime syllable in the Test Trial Types: (a) an Illegal consonant cluster (e.g., /dl/), (b) a cluster that is Absent in English (e.g., /bw/), or (c) a Legal but rare consonant cluster (e.g., /dw/). Prime Status conditions are referred to henceforth as Illegal, Absent, and Legal. The third factor was the voicing of the first consonant in both the prime and target syllable; this stop consonant was either voiced (e.g., /gw/) or voiceless (e.g., /kw/). The fourth factor was whether the prime and target were produced by the same speaker or different speakers. These third and fourth factors were shown not to interact with any other factors, so will not be discussed further.

Stimuli consisted of CCV syllables, in which the CC clusters were those just described and listed in Table 1, followed by one of the three vowels /a, e, i/. Two native speakers of Hebrew, who also speak English, recorded the stimuli in a sound-attenuating room. Each syllable was produced multiple times with 5 different intonations, corresponding to various emotional states (e.g., surprised, happy, sad). In this way, 15 tokens of each syllable type (5 intonations x 3 vowels) were recorded. The stimuli were spoken into a head-mounted microphone and digitized at a sampling rate of 44.1 kHz with 16-bit resolution.

The individual syllables were then cut out of the continuous recording using a waveform editor, down-sampled to 16 kHz, and equalized in peak intensity. From the initial recordings, five prime-target pairs were selected to exemplify each of the 36 conditions. Pairs were selected such that there was high similarity within the pairs (i.e., same intonation and vowel) and low similarity across the pairs (i.e., different intonation and vowel). To ensure that similarity ratings were not made based entirely on the basis of nonlinguistic characteristics, half of the trials included primes and targets recorded by different speakers. The final group of selected and edited syllables had a mean duration of 293 ms ($SD = 43$ ms). Comparison of the spectrograms and pitch contours in the left and right panels of Figure 1 demonstrates within-pair similarity; comparison of the spectrograms and pitch contours on the top and bottom panels demonstrates across-pair variability.

Procedure

Participants were seated in a comfortable chair located 150 cm from the computer monitor. Each trial began with the presentation of a fixation cross which remained on the computer monitor until the response cue appeared. The prime syllable began 2000 ms after the onset of the fixation point and the target stimulus began 1200 ms after the onset of the prime. Sounds were presented from 2 loudspeakers on either side of the monitor at an average intensity of 57 dBA measured at the location of participants. The response cue (“How similar? 1 = most similar; 4 = most dissimilar”) appeared 500 ms after the offset of the target and remained on the screen until the participant responded or 5 seconds elapsed. Short breaks were offered approximately every 15 minutes. Participants were encouraged to ask for longer breaks as needed. Each of the five prime-target pairs from the 36 conditions was

presented 10 times over the course of the experiment, resulting in a total of 1800 trials. These trials were presented in random order in 15 blocks of 120 trials each. The entire experiment session took 3–3.5 hours to complete.

Continuous EEG was sampled at 250 Hz and a bandwidth of .01–100 Hz throughout the duration of the experiment from 128-channel HydroCel Geodesic Sensor Nets (Electrical Geodesics Inc., Eugene, OR). Impedance was brought below 50 k Ω at every electrode at the beginning of the experiment and maintained below 100 k Ω for the duration. The continuous EEG was divided into 1000 ms epochs, from 100 ms before to 900 ms after the onset of a target. Trials that contained eye blinks, eye movements, or other identifiable artifacts were excluded from analysis by setting subject-specific amplitude thresholds based on visual inspection of the raw data. Data from the remaining trials were averaged by subject and condition, re-referenced to the average of the two mastoid electrodes, and corrected to the 100 ms pre-target baseline. Subjects included in the analysis contributed data from 83–198 ($M = 176$) out of a possible 200 trials in each condition.

Analysis

The similarity ratings were analyzed with mixed effect models of three binomially distributed dependent variables: the proportions of 1 versus 2–4, 1–2 versus 3–4, and 1–3 versus 4 ratings. These partitions simulate an ordinal logistic regression. The random effects were participants and items and the possible fixed effects Trial Type and Prime Status.

Data from 72 electrodes across the entire scalp were included in analysis such that waveforms averaged across six recording sites were entered into ANOVAs with four levels of Anterior-Posterior position and three levels of Laterality (left, medial, and right). Approximate locations of electrodes and electrode-position factors in the ANOVAs are included in the ERP waveform figures. Amplitude measurements were taken at each electrode in four time intervals: 50–80 ms (P1); 90–140 ms (N1); 200–350 ms, and 350–600 ms. Average amplitude in each time range was entered into a 3 (Trial Type) $\times$ 3 (Prime Status) $\times$ 4 (Anterior-Posterior position) $\times$ 3 (Laterality) Greenhouse-Geisser-adjusted repeated-measures ANOVA. All significant main effects and interactions ($p < .05$) were further investigated with post-hoc analyses.

RESULTS

Similarity ratings

Histograms of the similarity ratings, pooled across participants, items, and voicing but separated by Trial Type and Prime Status, are displayed in Figure 2. Ratings of 1 (i.e., most similar) predominate for all Identity trials regardless of Prime Status (top row: panels a, d, g), while ratings of 4 (i.e., most dissimilar) predominate for all Control trials regardless of Prime Status (bottom row: panels c, f, i). Ratings of 4 also predominate for Test trials when Prime Status is either Legal (panel b) or Absent (panel h). In contrast, for Test trials with an Illegal prime (panel e), the majority of ratings were 1s and 2s. These distributions indicate that on Test trials, the participants consistently considered Legal and Absent primes but not Illegal primes to differ from targets. This result is predicted from previous demonstrations of

English listeners' inability to distinguish /d, t/ from /g, k/ before /l/ (Hallé & Best, 2007; Hallé et al., 2003; Moreton, 2002).

The previous discrimination data further predict that the ratings for Illegal primes on Test trials should resemble those on Identity trials. In other words, there should be no difference in the extent to which the Test primes /dl, tl/ and Identity primes /gl, kl/ are considered to be similar to the targets /gl, kl/. This prediction was confirmed in the mixed effects models. The fixed effect Trial Type (Identity versus Test within the Illegal Status) was not significant for any of the three partitions of the response categories (1:2–4 $z = -1.00$, $p > 0.25$; 1–2:3–4 $z = -1.13$, $p > 0.25$; 1–3:4 $z = -0.85$, $p > 0.25$).

Moreton (2002) showed that English listeners do not disfavor “b” over “g” responses before /w/, despite the absence of /bw/ (and /pw/) clusters in English. This observation predicts that on Test trials in the current experiment the Absent primes /bw, pw/ and Legal primes /dw, tw/ should be considered to be equally different from the targets /gw, kw/. The mixed effect models also confirm this prediction. Prime Status (Absent versus Legal across Trial Types) was not significant for any partition of response categories (1:2–4 $z = 0.63$, $p > 0.50$; 1–2:3–4: $z = -0.69$, $p > 0.40$; 1–3:4: $z = -0.68$, $p > 0.40$).

ERPS Time-locked to Targets

Figure 3 shows the grand average ERPs elicited by target syllables for all Trial Types in the Legal and Illegal Prime Status conditions. Target sounds elicited a typical positive-negative-positive series of peaks that were largest over anterior medial regions. Targets elicited a positive deflection (P1) that peaked at 70 ms post-stimulus and a negative deflection (N1) between 90 and 140 ms that peaked at 103 ms. There were no differences in the amplitude or latency of these first two peaks for the experimental conditions (p 's > 0.50).

Mean amplitude 200–350 ms—Between 200 and 350 milliseconds after onset, targets elicited a positivity that varied in amplitude by condition, as shown in Figures 3 and 4. More priming is indicated by smaller positive deflections.

We first compared the most extreme Prime Status conditions, Legal and Illegal. Over anterior regions of the right hemisphere, there was a main effect of Trial Type (Identity, Control, Test), $F(2,34) = 8.30$, $p < .005$, in which targets that followed Control and Test primes elicited more positive ERPs than targets that followed Identity primes. This main effect was significant for both Legal primes, $F(2,34) = 4.08$, $p < .05$, and Illegal Primes, $F(2,34) = 5.70$, $p < .01$. Paired t-tests showed that ERPs for targets which followed Test primes were more positive than those that followed Identity primes, for both Legal primes, $t(17) = 2.95$, $p < .001$, and Illegal ones, $t(17) = 2.03$, $p = .06$. In contrast, the amplitude of this positivity did not differ significantly following the Test and Control primes for either Legal or Illegal primes, p 's $> .11$. Further, Trial Type did not interact with Prime Status (Legal, Illegal, $p = .18$). Analysis of data collected at right anterior electrodes was initially motivated by Trial Type and Prime Status interactions with electrode position factors when data across the entire scalp were included [Trial Type x Prime Status x Laterality: $F(4,68) = 2.44$, $p = .06$; Right electrodes only, Trial Type x Anterior- Posterior: $F(6,102) = 14.90$, $p < .001$].

ERPs elicited by targets in the Legal and Absent conditions were also compared, as shown in Figure 4. Over the same anterior, right-hemisphere regions there was a main effect of Trial Type, $F(2,34) = 12.70$, $p < .001$. As was true for the Legal and Illegal primes, for the Absent prime condition alone, there was a main effect of Trial Type, $F(2,34) = 12.70$, $p < .001$. Targets following Test primes elicited a larger positivity than those following Identity primes ($t(17) = 4.74$, $p < .001$); targets following Control primes also elicited a larger positivity than those following Identity primes ($t(17) = 3.52$, $p < .005$); and Test and Control conditions did not differ ($p > .21$). As in the Legal versus Illegal comparison, Trial Type and Prime Status did not interact for the Legal and Absent conditions ($p > .12$). The results of an ANOVA on measurements taken across the entire scalp again motivated analyzing data from the right anterior positions [Trial Type x Prime Status x Laterality: $F(4,68) = 3.80$, $p < .05$; Right electrodes only, Trial Type x Anterior-Posterior: $F(6,102) = 9.58$, $p < .0001$]. Mean amplitude from these electrodes of interest is shown in Figure 5.

Mean amplitude 350–600 ms—Unlike in the earlier time window, between 350 and 600 ms, there was a significant interaction of Trial Type and Prime Status (Legal, Illegal) on ERPs elicited by targets over posterior regions, $F(2,34) = 7.48$, $p < .005$, even though the main effects of Trial Type were significant following both Legal primes, $F(2,34) = 10.93$, $p < .005$, and Illegal primes, $F(2,34) = 8.85$, $p < .005$. The interaction was significant because for Legal Primes, ERPs elicited by targets did not differ following Control and Test primes, $p > .25$, but for Illegal primes, ERPs did not differ following Identity and Test primes, $p > .92$. Targets following Legal Identity primes elicited a larger positivity than those following both Test and Control primes, [Test, Identity: $t(17) = 4.60$, $p < .001$, Control, Identity: $t(17) = 3.12$, $p < .01$], but targets following both Illegal Identity and Test primes elicited a larger positivity than targets following Illegal Control primes [Control, Test: $t(17) = 3.49$, $p < .005$, Control, Identity: $t(17) = 3.20$, $p < .01$]. These results match the behavioral data: listeners rated targets as being equally similar to Illegal Identity and Test primes, and as being equally dissimilar from Legal Control and Test primes. Analyzing data from the posterior electrodes alone was motivated by the interaction of Trial Type and Prime Status with electrode position factors across the entire scalp [Trial Type x Anterior-Posterior x Laterality: $F(12,204) = 3.69$, $p < .01$; Prime Status x Anterior-Posterior x Laterality: $F(6,102) = 2.84$, $p = .06$].

The effects of Trial Type and Prime Status on mean amplitude between 350 and 600 ms over posterior regions were also examined for the Legal and Absent conditions, as shown in Figure 4. These factors interacted marginally on measurements over these regions, $F(2,34) = 2.80$, $p = .07$. For the Absent condition alone, there was main effect of Trial Type, $F(2,34) = 3.30$, $p = .06$. In contrast to the Legal condition, where Control and Test Trials both differed from Identity trials, differences in the Absent condition were driven by the difference between the Control and both Identity trials, $t(17) = 3.02$, $p < .01$, and Test trials, $t(17) = 1.93$, $p = .07$, which did not differ from one another, $p > .59$. Analyses that included measurements at posterior electrodes only were motivated by Trial Type by electrode position interactions across the scalp [Trial Type x Anterior-Posterior x Laterality: $F(12,204) = 4.94$, $p < .001$]. Mean amplitude from these electrodes is shown in Figure 6.

DISCUSSION

The current experiment was designed to determine *when* assimilation of non-native speech sequences takes place. Specifically, we tested the predictions of the autonomous and interactive hypotheses of speech perception. The former predicts that speech sounds within phonotactically illegal clusters are faithfully represented in early representations and that only later, and possibly more abstract, representations are repaired to fit native language patterns. In contrast, the interactive hypothesis predicts that this repair occurs much earlier such that more sensory perceptual representations are influenced by the phonotactic requirements of the listener's native language. In addition to testing the extremes of phonotactically frequent and illegal clusters, we also measured the effects of hearing phonotactically absent and rare primes on processing target syllables. The behavioral and ERP results support two main conclusions. First, the early representations of illegal sequences are veridical; only later representations are influenced by native language phonotactics patterns. Second, phoneme combinations that are legal but rare or absent are represented faithfully in both early and later processing.

Replicating prior work (Hallé & Best, 2007; Hallé et al., 2003; Hallé et al., 1998; Massaro & Cohen, 1983; Moreton, 2002), we demonstrated that English listeners explicitly report hearing illegal strings as legal syllables that differ by one phonological feature. Specifically, participants rated phonotactically illegal syllables /dla, tla/ as being highly similar to legal syllables which differed just in place of articulation /gla, kla/. In contrast, phonotactically legal control syllables that also differed just in voicing /kla, gla/ were rated as highly dissimilar. The same was not true for the absent-but-not-illegal syllables /bwa, pwa/ or the rare-but-legal syllables /dwa, twa/ which were all rated as dissimilar from targets that differed by place of articulation /gwa/, kwa/.

The ERP results indicate that hearing a phonotactically illegal cluster affects early processing of a subsequent legal cluster in a manner that is not captured by behavioral responses. Between 200 and 350 ms after the onset of the target /gla/, neither illegal primes /dla/ nor control primes /kla/ facilitate processing of the target to the same extent as different tokens of the same syllable /gla/. In contrast, between 350 and 600 ms after target onset, illegal and identity primes exert a similar influence over target processing. These results are consistent with the autonomous model of speech perception, which predicts that early representations of speech sounds are veridical and only later representations are regulated by language-specific phonotactic constraints. Importantly, the differences in the early representations of illegal clusters and the legal sequences they are reported to be similar to are not ephemeral; they continue to influence the early perceptual processing of other legal syllables presented 1200 ms later (which was the interval between the onsets of the prime and target syllable in this study).

Our results are inconsistent with those reported by Dehaene-Lambertz et al. (2000), who observed no differences in ERPs recorded from Japanese speakers when there was and was not a switch between legal (e.g., /igumo/) and illegal strings (e.g., /igmo/). There are at least three reasons why our results may have differed. First, we investigated a different phonotactic constraint. Participants in Dehaene-Lambertz et al.'s study had to detect whether

a sound had been added or subtracted, while participants in our study had to detect a change in a sound's feature, specifically its place of articulation. That their participants showed no evidence of representing the illegal sequence suggests, perhaps paradoxically, that a feature change is more salient than adding or subtracting an entire sound. Second, they compared results across participants, but we compared results within participants. Our within-subjects priming measures may have provided the additional power necessary to detect differences in the representations of the illegal clusters and their legal counterparts. Third, both the tasks and the ERP measures employed in the two studies were quite different, which could influence differences across the studies. In particular, the French and Japanese speakers may have differed in how engaged they were in detecting a change between standards and deviants. The inability of Japanese speakers to hear differences in any of the critical stimuli may have led them to stop forming detailed representations of the sounds. In the current experiment, subjects reported hearing differences between the prime and target more than 50% of the time.

The ERP results clearly distinguish between early, veridical representations and later ones influenced by language experience. Those early representations could consist of the auditory qualities evoked by the stimuli, and the representations of [d] or [t] and [g] or [k] would differ enough in their values for those auditory qualities before [l] that they can be distinguished there as easily as in other contexts. The later representations might consist instead of the feature values that distinguish one phoneme from another, such that representations of /d/ or /t/ and /g/ or /k/ would no longer differ before /l/ as they do in other contexts because the phonotactics prohibits the phonemes /d, t/ in that context. That is, representations consisting of auditory qualities are necessarily veridical, perhaps because they are produced by a pre-linguistic or linguistically naïve stage in processing, but those consisting of feature values are not if the grammar does not permit them to be. These characterizations of the content of the representations at the two stages in processing are compatible with our ERP data, but our results do not require that the contents of the representations differ at the two stages in processing. The early stage could consist instead of language-specific phones (cf. Norris et al., 2000) rather than auditory qualities, and the architecture of processing could delay the application of phonotactic constraints until later, perhaps because the evaluation of each phone is first context-free and only later context-sensitive.

One way to determine what the contents of the representations are that influence the early versus late ERPs would be to manipulate characteristics of the discrimination task. Fujisaki and Kawashima (1969, 1970) showed that listeners discriminated within-category stimuli significantly better than predicted by their categorization of those same stimuli. They interpreted this result as evidence that listeners could use two kinds of memory of the stimuli to discriminate them. Using auditory memory would permit listeners to discriminate physically different stimuli that belong to the same category, while relying on phonetic memory would instead limit them to discriminating stimuli that belonged to different categories. In line with this distinction, Pisoni (1973) demonstrated that listeners' ability to discriminate stimuli that they gave the same label to depended on the inter-stimulus interval, a result that represents the fading of auditory but not phonetic memory. The ISI between prime and target in our stimuli, 1200 ms, was, however, far longer than Pisoni found to be

sufficient for auditory memory to have faded, so the difference between Test and Identity Illegal pairs may not reflect listener's retention of a raw auditory representation of the prime. Other evidence that categorization does not fully account for discrimination can be found in studies by Pisoni and Lazarus (1974), Carney, Widen, and Viemeister, (1977), and Gerrits and Schouten (2004), who all showed that within-category discriminability depends on the format of the discrimination task itself, and in Pisoni and Tash (1974), who found that response times were systematically slower to stimuli near category boundaries than those near category prototypes. In a similar vein, McMurray, Aslin, Tanenhaus, Spivey, and Subik (2008) reported that eye movements to a competing word were more frequent for stimuli close to the category boundary than to those farther from it. Together, these results indicate that percepts grade continuously with stimulus properties rather than consisting of unanalyzable categories.

The behavioral studies cited above demonstrate that gradient perceptual similarity is reflected in gradient responses. We propose that the early ERP effects observed in the current study also reflect gradations in perceived acoustic similarity between prime and target. Specifically, the early positivity, between 200 and 350 ms, likely reflects the effects of acoustic priming, such that the more similar the target is to the prime, the smaller the positivity in this time window. These results are consistent with those of previous acoustic priming studies. For example, Coch et al. (2005) observed a larger early left anterior negativity (similar to our observed smaller positivity) to non-word targets preceded by rhyming primes (e.g. *hane* preceded by *blane*) than to targets preceded by non-rhyming primes (e.g. *lale* preceded by *crute*). And, similar to the current results, this effect changed polarity in later time windows, such that rhyming targets elicited a greater positivity in posterior sites starting 300 ms after target onset (see also Coch et al., 2002 who observed the same results with real word stimuli.)

As noted, this gradient priming effect is surprising given the long inter-stimulus interval (1200 ms) between prime and target in the current study. The fact that the acoustic similarity of prime to target affects early target processing after such a long delay suggests that listeners can after all store an accurate perceptual representation over the delay. Once again, this result is similar to that observed by Coch and colleagues (2002, 2005), who used a similar ISI of 1167 ms in their studies.

One alternative interpretation of the current results is that the early, reduced positivities elicited by targets are in fact instances of the Phonological Mapping Negativity (PMN), which is elicited between 200 and 350 ms after target onset when auditory input does not match phonological expectations (Connolly & Phillips, 1994; Desroches, Newman, & Joannis, 2009; Newman & Connolly, 2009). For example, Newman and Connolly (2009) demonstrated a larger PMN in cases where subjects imagined a word or nonword (e.g., *snap* without the /s/; *snoth* without the /s/) then heard a mismatching prompt (e.g., *tap*; *toth*) than when the prompt matched the imagined word or nonword. Desroches et al. (2009) demonstrated a PMN to targets preceded by rhymes (e.g., *bone* preceded by *cone*) compared to identical words (e.g., *cone* preceded by *cone*) but not to targets preceded by cohort pairs (e.g., *comb* preceded by *cone*). These results are interpreted as evidence that the PMN reflects a mismatch between top-down expectations and phonological input. However, the

current results are not consistent with a PMN interpretation, as the pattern of results we observe is, in fact, the opposite of previous manifestations of this component. Specifically, we would expect the largest PMN in the Control conditions, where the prime and target differ in one phonological feature (e.g., /gla/ preceded by /kla/). However, it is in these conditions where we observe the largest *positivity*. Therefore, it appears that the current results are not reflecting the storage of phonological expectations from the prime to the target.

The second ERP component of note is the late positivity observed between 350 and 600 ms after target onset, which we interpret as an example of a P3, or P300. The P3 is believed to index target detection (Sutton, Braren, Zubin, & John, 1965; Squires, Hillyard, & Lindsay, 1973; Donald & Little, 1981). Specifically, we observed the largest P3s to targets preceded by identical primes. Furthermore, P3 amplitude has been shown to increase with increasing confidence (Squires et al., 1973), which is consistent with our finding that the late positivity in the current study closely mirrors the behavioral results, such that the more similar the prime and target were rated, the larger the P3 we observed for the target.

In conclusion, the results of the current study are consistent with the autonomous hypothesis of speech perception, in which early representations of speech are veridical and only later representations are mapped onto language-specific categories. These results can not only inform the further specification of models of speech perception, but also offer an ERP method that can be used to explore additional questions about the perception of native and non-native speech sounds.

Acknowledgments

The research reported here was partially funded by R01-DC006241 awarded to J.K., a John Merck Fellowship in the Biology of Developmental Disabilities awarded to L.D.S., and NIH Institutional Training Grant MH16745, which provided postdoctoral training for M.B. The content is solely the responsibility of the authors and does not necessarily represent the official views of the John Merck Foundation, NIMH, or NIH. We thank Michael Becker and Shai Cohen for assistance with stimulus preparation, Martin Nolet and Paul Dainton for assistance with data collection and analysis, and Lori Astheimer, Patrick Taylor, Ben Zobel, Nick Planet, and Will Bush for comments on previous versions of the manuscript.

References

- Brown R, Hildum D. Expectancy and the perception of syllables. *Language*. 1956; 32:411–419.
- Carney A, Widin G, Viemeister N. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*. 1977; 62:961–970. [PubMed: 908791]
- Coch D, Grossi G, Coffey-Corina S, Holcomb P, Neville H. A developmental investigation of ERP auditory rhyming effects. *Developmental Science*. 2002; 5:467–489.
- Coch D, Grossi G, Skendzel W, Neville H. ERP nonword rhyming effects in children and adults. *Journal of Cognitive Neuroscience*. 2005; 17:168–182. [PubMed: 15701247]
- Connolly J, Phillips N. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*. 1994; 6:256–266. [PubMed: 23964975]
- Dehaene-Lambertz G, Dupoux E, Gout A. Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*. 2000; 12:635–647. [PubMed: 10936916]

- Desroches A, Newman R, Joannise M. Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*. 2009; 21:1893–1906. [PubMed: 18855555]
- Donald M, Little R. The analysis of stimulus probability inside and outside the focus of attention, as reflected by the auditory N1 and P3 components. *Canadian Journal of Psychology*. 1981; 35:175–187. [PubMed: 7248850]
- Dupoux E, Kakehi K, Hirose Y, Pallier C, Mehler J. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*. 1999; 25:1568–1578.
- Dupoux E, Pallier C, Kakehi K, Mehler J. New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*. 2001; 16:491–505.
- Fujisaki H, Kawashima T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*. 1969; 28:67–73.
- Fujisaki H, Kawashima T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute*. 1970; 29:207–214.
- Gerrits E, Schouten M. Categorical perception depends on the discrimination task. *Perception and Psychophysics*. 2004; 66:363–376. [PubMed: 15283062]
- Grossi G, Coch D, Coffey-Corina S, Holcomb P, Neville H. Phonological processing in visual rhyming: A developmental ERP study. *Journal of Cognitive Neuroscience*. 2001; 13:610–625. [PubMed: 11506660]
- Hallé P, Best C. Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America*. 2007; 121:2899–2914. [PubMed: 17550188]
- Hallé, P., Best, C., Bachrach, A. Perception of /dl/ and /tl/ clusters: A cross-linguistic perceptual study with French and Israeli listeners. *Proceedings of the 15th International Congress of Phonetic Sciences*; Barcelona. 2003. p. 2893-2896.
- Hallé P, Segui J, Frauenfelder U, Meunier C. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*. 1998; 24:592–608. [PubMed: 9554098]
- Massaro D, Cohen M. Phonological context in speech perception. *Perception and Psychophysics*. 1983; 34:338–348. [PubMed: 6657435]
- McClelland J, Elman J. The TRACE model of speech perception. *Cognitive Psychology*. 1986; 18:1–86. [PubMed: 3753912]
- McMurray B, Aslin R, Tanenhaus M, Spivey M, Subik D. Gradient sensitivity to within-category variation in speech: Implications for categorical perception. *Journal of Experimental Psychology, Human Perception and Performance*. 2008; 34:1609–1631. [PubMed: 19045996]
- Moreton E. Structural constraints in the perception of English stop-sonorant clusters. *Cognition*. 2002; 84:55–71. [PubMed: 12062147]
- Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. 2007; 118:2544–2590. [PubMed: 17931964]
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iovonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. 1997; 385:432–434. [PubMed: 9009189]
- Newman R, Connolly J. Electrophysiological markers of pre-lexical speech processing: Evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology*. 2009; 80:114–121. [PubMed: 18524453]
- Norris D, McQueen J, Cutler A. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*. 2000; 23:299–370. [PubMed: 11301575]
- Pisoni D. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*. 1973; 13:253–260. [PubMed: 23226880]
- Pisoni D, Lazarus J. Categorical and non-categorical modes of perception along the voicing continuum. *Journal of the Acoustical Society of America*. 1974; 55:328–334. [PubMed: 4821837]

- Pisoni D, Tash J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*. 1974; 15:285–290. [PubMed: 23226881]
- Schröger E. Mismatch negativity: A microphone into auditory memory. *Journal of Psychophysiology*. 2007; 21:138–146.
- Squires K, Hillyard S, Lindsay P. Vertex potentials evoked during auditory signal detection: Relation to decision criteria. *Perception & Psychophysics*. 1973; 14(2):265–272.
- Sussman E. A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology*. 2007; 21:164–175.
- Sutton S, Braren M, Zubin J, John E. Evoked-potential correlates of stimulus uncertainty. *Science*. 1965; 150:1187–1188. [PubMed: 5852977]
- Weber-Fox C, Spencer R, Cuadrado E, Smith A. Development of neural processes mediating rhyme judgments: Phonological and orthographic interactions. *Developmental Psychobiology*. 2003; 43:128–145. [PubMed: 12918092]
- Winkler I, Kujala T, Tiitinen H, Sivonen P, Alku P, Lehtokoski A, Czigler I, Csépe V, Ilmoniemi R, Näätänen R. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*. 1999; 36:638–642. [PubMed: 10442032]

Figure 1.

Spectrograms and pitch contours from two representative illegal prime-target pairs. On the left are illegal prime syllables, /dla/ and /dli/; on the right are the corresponding legal target syllables, /gla/ and /gli/. Comparison of primes and targets demonstrates the similarity of the syllables within a trial even on trials that included a Test or Control prime. Comparison of the spectrograms and pitch contours on the top and bottom panels demonstrates variability between the pairs presented on separate trials. All syllables shown here were produced by the same speaker.

Figure 2. Similarity ratings pooled across participants, voicing, and items, and shown separately by Trial Type and Prime Status.

Figure 3. Event-related potentials elicited by target syllables preceded by Identity primes (solid line), Control primes (dashed line), and Test primes (dotted line) in the (a) Legal (left) and (b) Illegal (right) Prime Status conditions. Shading indicates the time range and scalp regions at which statistically significant differences between at least two conditions were observed. The electrode map indicates the approximate locations where the data in the figure were recorded. Electrodes included in each level of Anterior/Posterior and Laterality in whole-head ANOVAs are also marked.

Figure 4.

Event-related potentials elicited by target syllables preceded by Identity primes (solid line), Control primes (dashed line), and Test primes (dotted line) in the (a) Legal (left, identical to Figure 3) and (b) Absent (right) Prime Status conditions. Shading indicates the time range and scalp regions at which statistically significant differences between at least two conditions were observed.

Figure 5. Average amplitude between 200 and 350 after target onset measured over right anterior regions. Error bars show standard error.

Figure 6.

Average amplitude between 350 and 600 after target onset measured over posterior regions. Error bars show standard errors.

Table 1

Consonant clusters from each of the 9 prime-target conditions, arranged by Trial Type (rows), Prime Status (columns), and Voicing (rows). Each cluster was followed by /a/, /e/, or /i/, but vowel was held constant within a prime-target pair within a trial.

Prime Status:	Legal		Illegal		Absent	
	Prime	Target	Prime	Target	Prime	Target
Identity Prime	Voiced	gw	gw	gl	gw	gw
	Unvoiced	kw	kl	kl	tw	tw
Control Prime	Voiced	kw	kl	gl	tw	gw
	Unvoiced	gw	gl	kl	gw	tw
Test Prime	Voiced	dw	dl	gl	bw	gw
	Unvoiced	tw	tl	kl	pw	tw