

Disentangling phonological well-formedness and attestedness: An ERP study of onset clusters in English

James White

Faith Chiu

University College London

Abstract

Disentangling the roles of phonological well-formedness and lexical attestedness in phonotactic processing has proven challenging. In this study, we present results from a passive listening ERP study showing that English speakers exhibit distinct neural responses to CCVC nonce words according to the phonological well-formedness and attestedness (in English) of the onset cluster. Clusters with poor sonority sequencing evoked an N400 effect compared to those without poor sonority sequencing, regardless of whether the well-formed clusters were attested in English. In contrast, unattested clusters, regardless of whether they were well-formed or ill-formed in terms of sonority sequencing, evoked a late positivity compared to attested clusters. The results suggest that listeners first perform a phonological analysis on potential words before submitting them to a lexical search.

Keywords: phonology, phonotactics, sonority, nonword acceptability, event related potentials

1. Introduction

It has long been noted that speakers have knowledge not only about the attested words of their language, but also about which hypothetical forms are phonologically possible words of their language. For instance, *brick* [brɪk] is both phonologically possible and attested as a word of English, *blick* [blɪk] is (until recently) unattested but deemed possible, and *bnick* [bnɪk] is both unattested and impossible (Chomsky & Halle 1965). Over the past several decades, a range of studies has demonstrated how rich and complex this phonotactic knowledge is. Speakers are able to access fine-tuned, gradient judgments about the acceptability of hypothetical words in their language (Frisch & Zawaydeh 2001; Hay et al. 2003; Albright 2009; Daland et al. 2011; Hayes & White 2013). Beyond acceptability judgments, phonotactic knowledge has been shown to affect performance in a variety of tasks, including the perception of illusory vowels (Dupoux et al. 1999; Moreton 2002; Davidson 2006, 2007; Berent et al. 2007), the parsing of ambiguous phrases (Hay et al. 2003), and the speed of nonce word repetition (Vitevitch & Luce 1998, 1999).

In this paper, we address the issue of whether phonotactic processing is solely dependent on the lexicon or is instead mediated by a phonological grammar. Phonotactic models diverge on this point. Some models assume that speakers generate an abstract phonotactic grammar from lexical statistics (perhaps filtered by other factors, as discussed below), and the grammar is then consulted whenever speakers make judgments about novel forms. For example, the Hayes and Wilson (2008) Phonotactic Learner uses the lexicon to create a grammar of weighted constraints, defined in terms of natural classes, that penalize certain sequences of sounds. Novel forms are tested by querying the grammar, not the lexicon

directly. However, exemplar models such as the Generalized Neighborhood Model (Bailey & Hahn 2001) evaluate the well-formedness of novel forms by comparing them directly to existing words in the lexicon. These accounts are not mutually exclusive; for instance, a novel form might be first compared to an abstract phonotactic grammar to determine whether it is well-formed, and then compared to the lexicon to determine its similarity to existing words.

If phonotactic processing occurs at two levels, once at a phonological level and once at a lexical level, then we would expect to find evidence of this bifurcation in psycholinguistic and neurolinguistic experiments. However, finding suitable phenomena for dissociating the two types of processing presents a challenge. A phonotactic grammar largely reflects the patterns in the lexicon; thus sequences missing from the lexicon are largely the same ones that we would expect to be judged as ill-formed by a grammar (see also discussion by Vitevitch & Luce 1998). What is needed is a case where a phonotactic grammar and the lexicon would make different predictions for the same set of forms.

In this study, we take the Sonority Sequencing Principle (SSP; Sievers 1881, Jespersen 1904, Hooper 1976, Steriade 1982, Selkirk 1984) as a test case. The SSP has been proposed as a universal phonological tendency whereby rises in sonority are preferred moving from the edges of a syllable to its nucleus. Under some accounts (e.g. see Clements 1992, Berent et al. 2007, Daland et al. 2011), the SSP represents a graded preference hierarchy: large rises > small rises > plateaus > small falls > large falls (where ‘>’ means ‘are more preferred than’). This graded preference hierarchy reflects implicational tendencies that we observe in typology; for instance, languages that allow small rises in onsets (e.g. [bn]) generally also allow large rises (e.g. [br]), but the reverse is not true.

Using a nonce word rating task, we first confirm that English speakers demonstrate both a sensitivity to the attestedness of a cluster and a gradient sensitivity to the SSP (for unattested clusters), replicating previous work by Daland et al. (2011). We then investigate the neural correlates of these behavioral results. Using a passive listening event-related potential (ERP) experiment (using electroencephalography, or EEG), we test whether speakers exhibit distinct neural responses for attestedness and for phonological well-formedness (i.e. SSP violations). Finally, we consider whether the ERP results reflect gradient processing of the SSP (consistent with the behavioral results) or categorical processing.

In the remainder of the introduction, we review potential sources of phonotactic knowledge, the SSP, and recent ERP studies looking at phonotactic processing before introducing our own study in more detail.

1.1. Sources of phonotactic knowledge and the SSP

A large body of literature has focused on understanding the source(s) of a speaker’s phonotactic knowledge, an issue that is not yet fully resolved. Much of a speaker’s phonotactic knowledge must be based on their lexicon. The phonotactic intuitions found in experiments often correspond to statistical properties of the lexicon, indicating that speakers can readily extract phonological generalizations from the attested words of their language (Coleman & Pierrehumbert 1997; Frisch & Zawaydeh 2001; Hay et al. 2003; Albright 2009; Daland et al. 2011; Hayes & White 2013). Since the types of words attested in the lexicon

vary greatly across languages, most of a speaker's phonotactic knowledge will be learned on a language-specific basis. These points strike us as uncontroversial.

However, several studies suggest that a speaker's phonotactic knowledge is not based solely on the lexical statistics. For instance, Hayes & White (2013) tested whether English speakers were influenced by phonologically natural and unnatural phonotactic constraints in a nonce word acceptability task. Both types of constraints were assigned high weights by the Hayes & Wilson (2008) Phonotactic Learner because they represented true statistical gaps in the English lexicon. However, the natural constraints had a strong effect on acceptability ratings whereas the unnatural constraints had little to no effect. Similar results have been found in Hungarian (Hayes et al. 2009) and Turkish (Becker et al. 2011). These studies suggest that lexical statistics are more likely to be extrapolated as phonotactic generalizations if they are consistent with principles of phonological naturalness. They also support the view that speakers query a grammar to make phonotactic judgments, rather than (solely) depending on the existing lexicon. Otherwise, it is unclear why some lexical gaps would affect judgments and others would not.

Another striking finding is that speakers make principled distinctions that are consistent with the SSP, even for sequences that they have never encountered in their own language. For instance, English speakers reliably judge sequences with falling sonority (e.g. [lb]) to be worse than those with a sonority plateau (e.g., [bd]), and a plateau to be worse than a small rise (e.g., [bn]), even though none of these sequences are attested in English (Daland et al. 2011). Similar distinctions have been found in production and perception (Davidson 2006, 2007; Berent et al. 2007); for instance, English speakers are more likely to hear an illusory schwa in a nonce word like [lbɪf] (repaired to [ləbɪf]) than in [bdɪf], even though they have no experience listening to words beginning with either [lb] or [bd] (Berent et al. 2007).

The explanation for this behavior remains controversial. Phonotactic models that depend solely on the frequency of segment sequences in the lexicon (e.g. the Phonotactic Probability Calculator of Vitevich & Luce 2004) are unable to account for the effect because the unattested sequences all have a lexical frequency of zero (Berent et al. 2007; Daland et al. 2011). If the SSP indeed cannot be derived from the lexical statistics, then its source must be universal biases rather than language-specific knowledge. However, Daland et al. (2011) show that the SSP might be learnable from the lexicon of English, even as it applies to unattested sequences, provided that the phonotactic learning model contains syllable structure and the ability to generalize over features (see also Albright 2009, Hayes 2011). Still, recent work by Jarosz and Rysling (2017) suggests that the intuitions of Polish speakers in a similar task cannot be fully explained by lexical statistics, even with more nuanced phonotactic models. The answer seeming to emerge is that the SSP arises from an interaction of lexical projection and universal phonological biases (or perhaps perceptual biases; we return to this in the discussion).

Determining *when* the SSP affects processing is a slightly different, but related, issue. For tasks where participants have time to think and give responses, such as most nonce word rating tasks, it is possible that the gradient sensitivity to the SSP stems from a sort of meta-linguistic reanalysis rather than from the initial processing of the stimulus. In the current study, we measure ERPs during passive listening. This allows us to capture online neural responses that emerge upon hearing the various clusters; in addition, the responses within a passive listening task are automatic. Both of these aspects allow us to capture participant responses before they have time to engage in meta-linguistic analysis.

1.2. ERP studies of phonotactic processing

Until recently, there were few studies focusing on the neural correlates of phonotactic processing, but the last decade has seen an uptick. Many of these studies have focused on the differences between attested and unattested sequences of sounds in a language. For instance, Rossi et al. (2011) found that German speakers exhibited a greater N400 after hearing CCVC nonce words with an unattested onset compared to nonce words with an attested onset. The N400 is a negative neural response occurring around 400 ms after the presentation of a triggering stimulus (Kutas & Hillyard 1980); it has been associated with several aspects of linguistic and non-linguistic processing depending on the nature of the task being performed (see Kutas & Federmeier 2011 for a review). In Rossi et al., the greater N400 for attested onsets was interpreted to mean that hearing nonce words induces a stronger lexical search if the word is phonotactically legal in the speaker's language. The EEG measurements in their experiment were taken during passive listening, suggesting that this application of phonotactic knowledge before lexical access is fairly automatic. Rossi et al. (2013) replicated this finding and showed that the N400 for legal words decreased after repeated exposure, similar to what happens with real words. However, as in behavioral studies, focusing just on the lexical attestedness of sequences of sounds makes it difficult to determine whether the observed effects were due to phonological well-formedness *per se*, or to similarity with other words in the lexicon.

Domahs et al. (2009) tested for neural correlates of Obligatory Contour Principle (OCP) violations. OCP violations occur when identical segments (or similar segments under some interpretations) are adjacent to one another (e.g., McCarthy 1986; Frisch 1996; Frisch et al. 2004).¹ Domahs et al. compared three categories of SCVC words in German: existing words (e.g. [ʃpek]), well-formed nonce words (e.g. [ʃpɛf]), and OCP-violating nonce words having identical second and third consonants (e.g. [ʃpɛp]). They found that both types of nonce words resulted in a greater N400 compared to existing words. Phonological well-formedness resulted in a different effect: the OCP-violating nonce words resulted in a Late Positive Component (LPC) compared to the well-formed nonce words.

Ulbrich et al. (2016) further investigated the distinction between phonological well-formedness and attestedness, focusing on the SSP in coda clusters. They compared four groups of nonce words by manipulating both attestedness of the coda cluster in German and well-formedness of the cluster according to the SSP. SSP-violating coda clusters were found to exhibit a greater N400 compared to SSP-conforming clusters. In the later LPC window, coda clusters that were both SSP-violating and unattested exhibited the strongest LPC effect. Like Domahs et al. (2009), Ulbrich et al. (2016) found distinct effects for attestedness and well-formedness, but in a slightly different pattern. Whereas Domahs et al. found earlier processing for attestedness (N400) compared to phonological well-formedness (LPC), Ulbrich et al. (2016) found an early effect of phonological well-formedness (N400). The authors suggested that these differences could be due to the types of stimuli presented: Domahs et al. (2009) included real German words in addition to nonce words whereas Ulbrich et al. (2016) included only nonce words. Following up on Ulbrich et al. (2016), Wiese et al. (2017) conducted a very similar study on Polish speakers, whose language allows

¹ Adjacency is interpreted here to mean adjacent on the same tier (as in autosegmental phonology; Goldsmith 1979), so two consonants are considered adjacent (on their tier) even if there is a vowel between them.

a wide range of SSP-violating clusters. The experiment once again revealed interacting effects for both attestedness and well-formedness, with a larger effect of well-formedness in the earlier window (N400) and a larger effect of attestedness in the later window (LPC); however, the overall pattern of results was more complex.

There has been little research on ERPs associated with phonotactics at time windows earlier than the N400 effect, particularly involving spoken words (as opposed to reading). Earlier effects are more closely associated with perception and attention-based responses than higher level linguistic processing (for reviews, see Kutas & Van Petten 1994, Key et al. 2005, Coulson 2007). However, Rossi et al. (2013) found that attested onset clusters in German elicited a greater positivity at around 240 ms (P2) compared to unattested clusters. Moreover, Hunter (2013) found that high probability words (i.e. those with greater phonotactic probability, neighborhood density, and cohort size) resulted in a greater positivity compared to low probability words at around 220 ms, though this positivity was delayed to around 400 ms when forms were nonce words (see also Hunter 2016).²

In sum, recent studies have suggested that there are distinct (yet interacting) processing effects for phonological well-formedness and attestedness. However, questions remain about when the phonological processing occurs and how the two effects interact. Which factors determine whether we see phonological well-formedness effects early (N400 or earlier) or late (LPC)? More research is needed to elucidate these issues.

1.3. Current study

In the current study, we contribute to the literature in this area by testing onset clusters with English speakers, collecting both ERP data and behavioral data.

Our primary goal was to test for distinct effects of onset attestedness in English and onset well-formedness. The list of onsets, taken from Daland et al. (2011), included attested onsets, unattested onsets, and marginally attested onsets. The unattested onsets also varied in terms of their sonority profile, ranging from large sonority rises to large sonority falls. For the ERP study, we divided the onsets into three groups for comparison: (1) Attested Well-formed, (2) Unattested Well-formed, and (3) Unattested Ill-formed. To test for an effect of attestedness, we compared the Attested Well-formed onsets to the Unattested Well-formed onsets. To test for an effect of phonological well-formedness, we compared the Unattested Well-formed onsets to the Unattested Ill-formed onsets. Following previous studies, we focused on two time windows, one around 450ms (N400) and one around 600ms (LPC). Based on Ulbrich et al.'s (2016) study of German coda clusters, we expected to find an earlier effect (N400) corresponding to phonological well-formedness and a later effect (LPC) corresponding to attestedness, possibly interacting with well-formedness. We also considered whether the effect of well-formedness was a gradient effect following the SSP by further subdividing the Attested Ill-formed group of clusters into sonority falls and sonority plateaus. If SSP violations are processed in a gradient manner, then we would expect a larger response to SSP-violating clusters with falls than to those with plateaus. An equal response for falls and plateaus would instead be more consistent with categorical processing, suggesting that SSP

² We did not have specific hypotheses about time windows earlier than N400 in the current study given the limited prior literature and the fact that our study involved all nonce words. However, we briefly consider this possibility in the discussion section.

violations are treated equally regardless of sonority profile (at least at this stage of processing).

For the behavioral task, we aimed to replicate the results of Daland et al. (2011) with our own participants, stimuli, and method. In addition to the value of replication generally, we considered a replication to be particularly important in this case. First, we considered it important to show that our participants displayed the same well-formedness intuitions for unattested onsets that have been reported in the literature. Second, we wanted to ensure that the intuitions held for our list of stimuli, which were short CCVC words with relatively large lexical neighborhoods (compared to CCVCVC words in Daland et al. 2011). Finally, our ERP task involved auditory-only presentation of the stimuli (unlike Daland et al. 2011 who used only orthographic presentation), so we wanted to ensure that the acceptability intuitions remained even when we withheld the orthography.

2. Method

2.1. Participants

Ten monolingual speakers of Standard Southern British English (mean age = 18) participated in the experiment. All participants were right-handed and reported no history of speech and hearing problems, and no history of psychiatric and neurological disorders. Participants received a small amount of monetary compensation or course credit for their participation.

2.2. Materials and apparatus

The stimuli consisted of 96 CCVC nonwords (henceforth just “words”). We used all of the 48 CC onsets from Daland et al. 2011. Each onset was combined with two of six possible VC rimes such that each rime was used an equal number of times.³ The onsets were divided into three groups according to their Attestedness in English: Attested onsets are used frequently in English, Marginal onsets occur rarely or in loanwords only, and Unattested onsets virtually never occur. The Unattested onsets were further divided into groups according to their Sonority Profile, ranging from -3 (large sonority falls) to 3 (large sonority rises) with 0 indicating a sonority plateau. The categories for both Attestedness and Sonority Profile were taken directly from Daland et al. 2011. A summary of the onsets and rimes used in the current study is given in Table 1. A full list of stimuli is given in the Appendix.

The auditory stimuli were recorded by a phonetically trained native speaker of Standard Southern British English. The recording was done in a soundproof booth with a RØDE NT1-A condenser cardioid microphone and an audio interface recorder (RME Fireface UC) using the software ProRec (version 2.3) implemented on Speech Filing System (Huckvale 2016) at a sampling rate of 44.1 kHz (16-bit samples per second). Several tokens of each word were recorded, and the token judged by the authors to be most faithful to the intended sequence was selected. Spectrograms of the stimuli were visually inspected to ensure that the consonants of the onset clusters were produced accurately, without an intervening schwa. The root mean square (RMS) amplitude of the stimuli was normalized prior to presentation.

³ Daland et al. (2011) combined the CC onsets with VCVC tails, creating CCVCVC nonce words as stimuli, so that the stimuli would have few lexical neighbors. In the current study, we opted for shorter VC rimes to shorten the material following the crucial onset and reduce unnecessary noise in the EEG recordings.

Table 1. List of CC onsets and VC rimes used in the experiment in orthography and IPA transcription. For unattested onsets, sonority profile is given, ranging from -3 to 3. Onsets and sonority profiles are taken from Daland et al. (2011).

Attested onsets			Marginal onsets		Unattested onsets by sonority profile	VC rimes
<i>tw</i> [tw]	<i>tr</i> [tɹ]	<i>sw</i> [sw]	<i>gw</i> [gw]	<i>shl</i> [ʃl]	3: <i>pw</i> [pw], <i>zr</i> [zɹ]	- <i>ace</i> [eis]
<i>shr</i> [ʃɹ]	<i>pr</i> [pɹ]	<i>pl</i> [pl]	<i>vw</i> [vw]	<i>shw</i> [ʃw]	2: <i>mr</i> [mɹ], <i>tl</i> [tɹ]	- <i>ike</i> [aik]
<i>qu</i> [kw]	<i>cr</i> [kɹ]	<i>cl</i> [kl]	<i>shn</i> [ʃn]	<i>shm</i> [ʃm]	1: <i>dn</i> [dn], <i>cm</i> [km], <i>fn</i> [fn], <i>ml</i> [ml], <i>nl</i> [nl]	- <i>ip</i> [ɪp]
<i>gr</i> [gɹ]	<i>gl</i> [gl]	<i>fr</i> [fɹ]	<i>vl</i> [vl]	<i>bw</i> [bw]	0: <i>dg</i> [dg], <i>pk</i> [pk]	- <i>eak</i> [ik]
<i>fl</i> [fl]	<i>dr</i> [dɹ]	<i>br</i> [bɹ]	<i>dw</i> [dw]	<i>fw</i> [fw]	-1: <i>lm</i> [lm], <i>ln</i> [ln], <i>rl</i> [ɹl]	- <i>ome</i> [əʊm]
<i>bl</i> [bl]	<i>sn</i> [sn]	<i>sm</i> [sm]	<i>vr</i> [vɹ]	<i>thw</i> [θw]	-2: <i>lt</i> [lt], <i>rn</i> [ɹn]	- <i>oon</i> [un]
					-3: <i>rd</i> [ɹd], <i>rg</i> [ɹg]	

All stimuli were presented during the passive listening ERP recording, but only a subset of the words were analyzed. Specifically, the onsets were grouped as follows for the ERP analysis: Attested Well-formed {*bl*, *br*, *cl*, *cr*, *dr*, *fl*, *fr*, *gl*, *gr*, *qu*, *pl*, *pr*, *tr*, *tw*}; Unattested Well-formed {*bw*, *fw*, *dw*, *gw*, *pw*, *thw*, *tl*, *vl*, *vr*, *vw*}; Unattested Ill-formed {*dg*, *dn*, *fn*, *cm*, *lm*, *ln*, *lt*, *ml*, *mr*, *nl*, *pk*, *rd*, *rg*, *rl*, *rn*}. Well-formed onsets were defined as an obstruent followed by an approximant (i.e. those with a large sonority rise); the other clusters, having less than ideal sonority, were categorized as ill-formed.⁴ Marginally attested onsets were grouped with the Unattested onsets for the purposes of the ERP analysis.

EEG data were collected using a 64-channel Biosemi Active Two system with four additional external electrodes (on the left and right mastoids, and at the right canthus and left cheek bone for horizontal and vertical EOG, respectively, to monitor eye movements). Recordings were made with a sampling rate of 2048 Hz. Audio stimuli were delivered via Etymotic ER-1 insert earphones.

2.3. Procedure

Participants were seated in a sound attenuated booth in front of a computer screen while wearing insert earphones. They were first familiarized to each of the stimuli by watching a PowerPoint presentation play on the screen. Each slide contained one word written in orthography in large black font in the centre of the slide. 500 ms after the written word appeared, the sound file was played through the insert earphones. After 3 sec, the next slide appeared with another word. Participants were told that they would be seeing and hearing several imaginary words of English. They were asked to pay attention to the words, but that they did not need to memorize them. They were also told that each word had only one syllable. All 96 words were presented once. Each participant saw the words in the same order, which was randomized before creating the slideshow.

After the initial familiarization, participants completed the pre-EEG rating task. They were instructed that they would be rating each imaginary word on a scale from 1 to 8 according to how likely they believed the word was to become a real word of English. A

⁴ We did not include clusters containing sibilant fricatives in the ERP analysis, as these have been argued by many to be structurally different than other clusters (Steriade 1982; van der Hulst 1984; Kaye 1992; Vaux 2004; Goad 2012).

rating of 1 was described as “so strange that it would be very unlikely for this word to be used as a new word of English,” and a rating of 8 was described as “so English-like that you would not be surprised to see it used as the name of a new product, company, website, or slang term.” In each trial, participants saw a word written on the screen in its orthographic form and then heard it played through the insert earphones 500 ms later. After hearing the word, participants rated the word by clicking one of the buttons marked with numbers 1 through 8. “Not English-like at all” appeared next to the button marked 1, and “Very English-like” appeared next to the button marked 8 as a reminder. Participants could replay each sound file once before making their choice by clicking a button marked “Replay.” The order of the stimuli was randomized anew for each participant. The rating task was implemented using Praat (Boersma & Weenink 2016).

During the passive listening task, participants remained seated while auditory stimuli were presented binaurally with an interstimulus interval of 1s, with random jitter added to the timing to reduce predictability. Stimulus order was randomized except that immediate repetitions of the same token were not allowed. There were a total of 3,840 trials (40 trials for each of the 96 words) separated over 4 recording blocks (960 trials per block). Each block lasted about 25 minutes, and participants were given a break between blocks. Presentation of the stimuli in this phase was purely auditory (orthography was not presented). During the task, participants watched cartoons on a screen with the volume muted. Participants were instructed that the words that they would be hearing were the same words that they had previously seen. Prior to recording, subjects were told to pay attention to the cartoon and that they did not need to pay attention to the words being played in the background. Subjects were also briefed regarding the EEG data collection protocol. They were asked to stay awake and to minimise movement and blinking so as to prevent adding unnecessary artefacts to the recording.

After the EEG recording, participants completed the post-EEG rating task. This task was identical to the pre-EEG rating task, except that the stimuli were presented in audio only; the words did not appear in orthographic form on the screen.

Stimuli in all parts of the experiment (i.e. the familiarization phase, EEG recording, and rating tasks) were presented at 60 dB/SPL. Subjects were given the opportunity to change the stimulus loudness at the beginning of the experiment but all subjects declined this option.

3. Results

3.1. Rating results

Figure 1 (left panel) shows the means scores for each onset cluster from the pre-EEG rating task (which included both orthographic and auditory presentation) plotted against the mean scores from the post-EEG rating task (which included only auditory presentation). Overall mean ratings were higher in the post-EEG task (4.81) than in the pre-EEG task (4.07), perhaps due to the repeated exposure to the items during the passive listening task. However, there was a very strong correlation between the pre-EEG and post-EEG ratings ($r = .92$), indicating that participants were fairly consistent in their relative ratings between clusters in the two tasks. One cluster appears to be an outlier (see Figure 1): [rn] was rated much more highly in the post-EEG task (5.22) compared to the pre-EEG task (1.39), suggesting that it may have been misperceived when participants did not have access to the orthographic form.

Below we focus on the post-EEG ratings, which involved auditory-only presentation like the passive listening task.

Overall, there was a strong correlation between the post-EEG ratings and the ratings reported in Daland et al. 2011 ($r = .89$). Figure 1 (right panel) shows a plot of the ratings in the two studies. We further analyzed the post-EEG rating results using mixed-effects linear regression models implemented in R (R Core Team 2016) using the *lme4* package (Bates et al. 2015). To test the effect of Attestedness, we fitted a linear mixed effects model to the numerical ratings in the post-EEG data. The model contained a fixed effect for Attestedness, coded into three categorical levels: Marginal (coded as the reference category), Attested, and Unattested. The model's random effects structure included random intercepts for Subjects and Items and by-Subject random slopes for Attestedness. The effect of Attestedness was significant according to a likelihood ratio test comparing the full model to a simpler model with the effect removed (implemented using the *anova()* function), $\chi^2(2) = 20.69$, $p < .001$. Ratings were significantly higher for Attested onsets ($M=6.35$, $SEM=.57$) than for Marginal onsets ($M=4.74$, $SEM=.69$); $\beta = 1.62$, $t = 5.16$. Ratings were also significantly higher for Marginal onsets than for Unattested onsets ($M=3.31$, $SEM=.75$); $\beta = 1.43$, $t = 3.24$.

Figure 2 shows the post-EEG ratings for Unattested clusters according to their Sonority Profile, ranging from -3 (large falls) to 3 (large rises). As the plot illustrates, there is a positive linear relationship between Sonority Profiles and ratings in the Unattested clusters ($r = .53$). To evaluate the significance of this effect, we fitted a new linear mixed effects model to the ratings for only Unattested clusters. The model contained a fixed effect for Sonority Profile (ranging from -3 to 3), as well as random intercepts for Subjects and Items and by-Subject random slopes for Sonority Profile. The effect of Sonority Profile was significant according to a likelihood ratio test comparing the full model to a simpler model with the effect removed, $\chi^2(1) = 4.24$, $p = .04$. The model estimates that each increase of 1 in the Sonority Profile corresponds to a statistically significant increase of $.27$ in the rating ($t = 2.13$).⁵

⁵ The cluster [rn], which we believe was likely misperceived, appears once again to be an outlier in Figure 2. Despite this, we report analyses with the cluster included to avoid cherry-picking the data. If [rn] is instead excluded, the relationship between sonority and rating is even stronger ($r = .81$) and the sonority effect remains significant according to the likelihood ratio test ($p < .01$).

Figure 1. Left panel: Mean ratings from the pre-EEG rating task and the post-EEG rating task. Right panel: Mean ratings from the Post-EEG rating task and Daland et al. (2011). Attestedness group is represented by shade of grey.

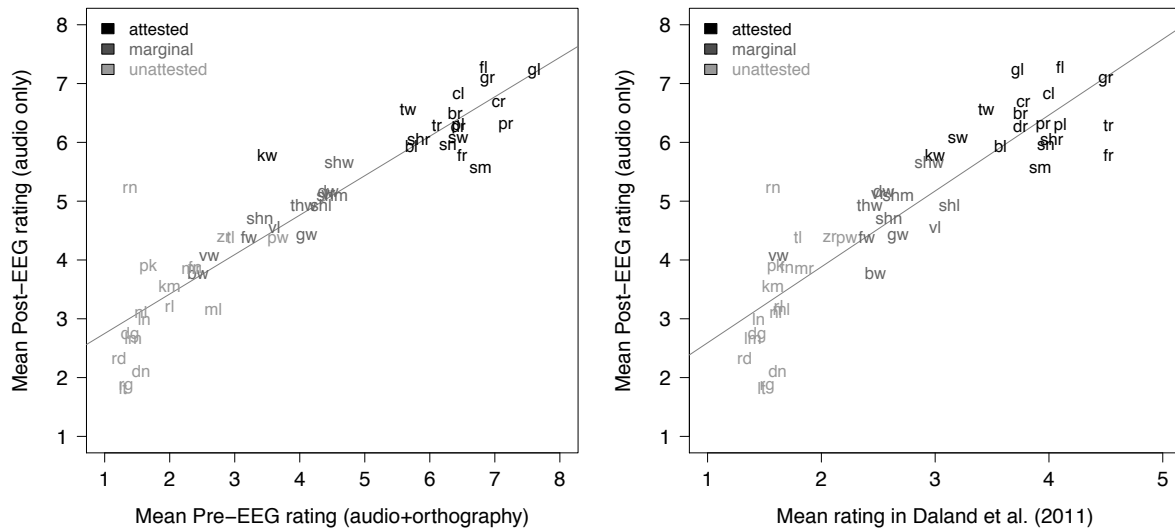
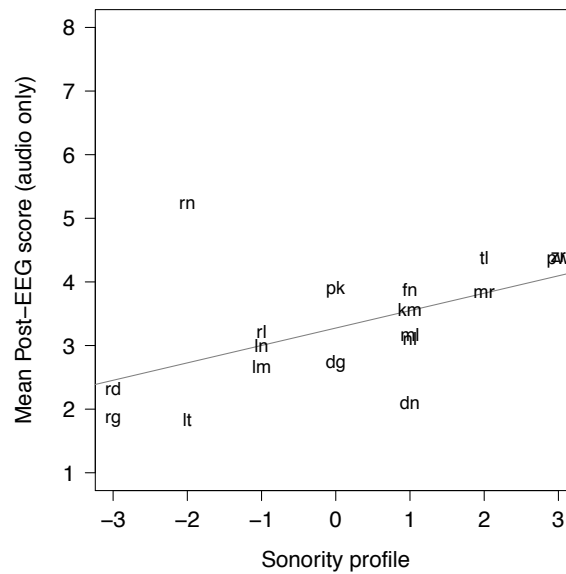


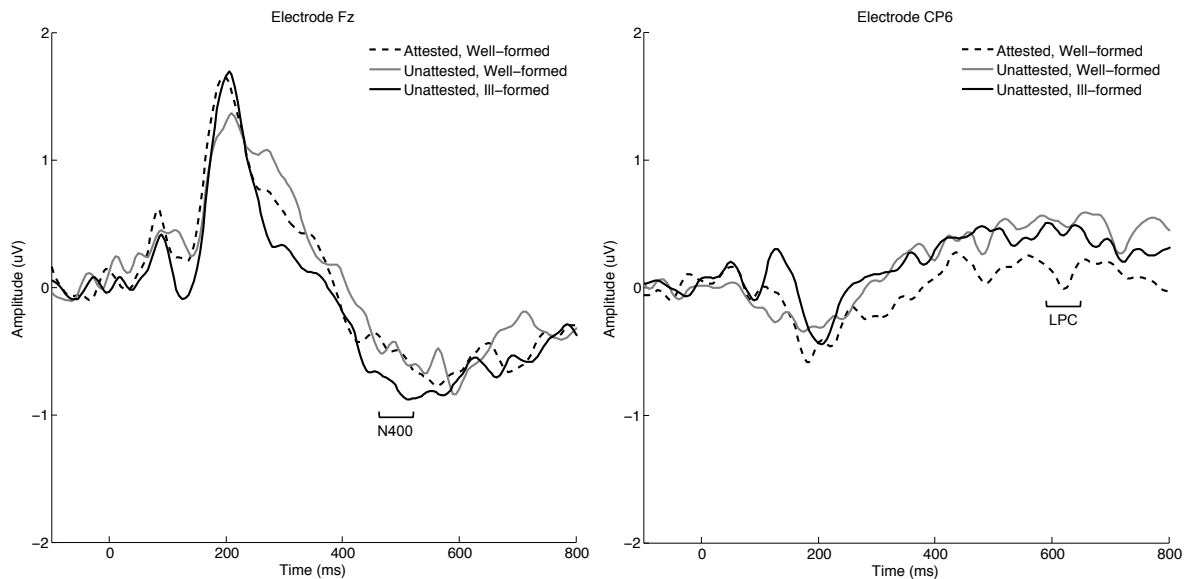
Figure 2. Mean ratings for unattested clusters in the post-EEG rating task according to sonority profile.



3.2. ERP results

EEG data were pre-processed on SPM12 (Ashburner et al. 2012) by reference to an average of all 64 scalp electrodes. A high-pass Butterworth filter was applied at 0.1Hz. Trials were epoched 100ms pre-stimulus and 900ms post-stimulus; the baseline average from the pre-stimulus interval was subtracted from each trial. Artefact removal was performed by thresholding all trials at $\pm 100\mu\text{V}$; interpolation was performed for any faulty channels. Trials not rejected from artefact removal were averaged per cluster (with both rimes) per subject, before further grouping into the analyzed conditions (Attested Well-formed, Unattested Well-formed, and Unattested Ill-formed).

Figure 3. Grand mean cortical auditory-evoked ERPs over 10 subjects. Left panel: Electrode Fz. Right panel: Electrode CP6.



In the following analysis, the dependent variable was the mean amplitude measured from two time windows (470–500ms and 590–620ms) placed around the maximum amplitude determined from visually inspecting grand-average waveforms across all channels. The mean amplitudes at 3 chosen electrodes were statistically analysed using repeated measures analysis of variance (ANOVA), with three within-subject factors: Time Window (470–500 ms and 590–620 ms), Stimulus Type (Attested Well-formed, Unattested Well-formed, and Unattested Ill-formed) and Electrode (Fz, CP5, and CP6). Electrodes chosen for statistical analysis are indicative of frontal-central (Fz) and bilateral posterior-central (CP5, left, and CP6, right) scalp areas. Mauchly’s Test indicated that the assumption of sphericity had not been violated for Stimulus Type ($\chi^2(2) = 3.377$, $p = .185$), nor the interaction between Electrode \times Stimulus Type ($\chi^2(9) = 11.027$, $p = .285$) or Electrode \times Time Window ($\chi^2(2) = 1.802$, $p = .406$). In cases where sphericity had been violated (Electrode main effect, and interactions Stimulus Type \times Time Window and Electrode \times Stimulus Type \times Time Window), Greenhouse-Geisser corrected p-values will be reported.

The results of a $2 \times 3 \times 3$ repeated measures ANOVA indicated a statistically significant main effect of Time Window ($F(1, 9)=35.211$, $p=.0001$, $\eta_p^2=.796$) due to selecting substantially different Event-Related Responses of different polarity, with the response around 470 ms being more negative in amplitude ($M=-.028$, $SEM=.056$) than the response around 600 ms ($M=.211$, $SEM=.050$). Neither the main effect of Electrode nor of Stimulus Type was significant. All interactions were significant, between Electrode \times Stimulus Type ($F(4, 36)=13.834$ $p=.001$, $\eta_p^2=.606$), Electrode \times Time Window ($F(2, 18)=13.978$ $p=.001$, $\eta_p^2=.608$), Stimulus Type \times Time Window ($F(2, 18)=10.276$ $p=.005$, $\eta_p^2=.533$), and Electrode \times Stimulus Type \times Time Window ($F(4, 36)=12.295$ $p=.001$, $\eta_p^2=.577$).

We followed up on the significant Electrode \times Stimulus Type \times Time Window interaction with specific comparisons of interest using separate paired samples t-tests. To examine a possible effect of well-formedness, we compared Unattested Well-formed onsets

and Unattested Ill-formed onsets at electrode Fz at the time window beginning 470ms. The mean amplitude of Unattested Ill-formed responses ($M=-.913$, $SEM=.356$) was significantly more negative than Unattested Well-formed responses ($M=-.466$, $SEM=.321$), $t(1,9)=2.395$, $p=.04$, see Figure 3 (left panel). The comparison between Attested, Well-formed clusters and Unattested, Well-formed clusters at the 600 ms time window revealed a near-significant effect of Attestedness at electrode CP6, $t(1,9)=-2.126$, $p=.06$. Unattested Well-formed clusters elicited a higher amplitude response ($M=.514$, $SEM=.243$) compared to Attested Well-formed clusters ($M=.0908$, $SEM=.24$), see Figure 3 (right panel).

To test whether there was a gradient response to SSP-violating clusters, we divided the Unattested Ill-formed clusters further into sonority plateaus and sonority falls. We conducted a paired samples t-test comparing responses at electrode Fz at the time window from 470 ms (i.e. the same time window and electrode where the general well-formedness effect was found). A paired samples t-test found no significant difference in response to clusters with sonority plateaus and those with sonority falls, $t(1,9) = .862$, $p=.41$.

4. Discussion

In this study, we conducted a behavioral task and a passive-listening ERP task to examine the relation between attestedness and well-formedness. In the behavioral task, we tested whether our participants were sensitive to both the attestedness of onsets in English and the sonority profile of unattested onsets; to this end, we aimed to replicate the findings of Daland et al. (2011) with our own stimuli and participants. In the ERP task, we tested whether English speakers would exhibit distinct neural responses to the attestedness and phonological well-formedness of onset clusters when passively listening to nonce words. Below, we discuss these issues in turn.

4.1. Nonce word rating results

Our nonce word acceptability results indeed showed that participants were sensitive both to the attestedness of onsets in English, and to the sonority profile of onsets that were unattested in English. The attestedness effect was realized as large differences in ratings between attested, marginal, and unattested onsets. The sonority effect was realized as a gradual improvement in ratings for unattested onsets as their sonority profile changed from -3 (large falls) to 3 (large rises), consistent with the SSP (see Figure 2). The gradual improvement suggests that participants were able to access gradient judgments about sonority sequencing even for onsets that are unattested in their language, consistent with previous research (Berent et al. 2007; Daland et al. 2011). Overall, our behavioral results can be considered a full replication of the findings reported by Daland et al. (2011) with different participants (who speak a different variety of English), different types of stimuli (simpler CCVC nonce words rather than CCVCVC nonce words), and a different manner of presentation (auditory rather than orthographic). Because our primary goal was to test for neural correlates of attestedness and well-formedness, it was important to confirm that our participants were sensitive to both factors in a behavioral task using the same stimuli that we used for collecting EEG measurements.

It is worth noting that we observed better differentiation of the unattested clusters in the post-EEG rating task than in the pre-EEG rating task (observable in Figure 1 as the slight curve in the linear trend on the left side of the plot). This could be due to the additional

exposure that participants received during the EEG recording; indeed, the ratings were slightly higher overall in the post-EEG rating task, which is likely an exposure effect. It could also be due to the difference in presentation: perhaps the sonority effect is stronger when not obscured by orthotactic effects stemming from presentation of the orthographic forms. We also found better differentiation of the unattested onsets than Daland et al. (2011) found in their rating task (observable in the right plot in Figure 1).⁶ A likely explanation for this is that Daland et al. (2011) used a 5-point scale whereas we used an 8-point scale, which allowed participants more opportunity to fully express subtle distinctions in their intuitions. The increased amount of exposure and the lack of orthography may have also contributed to the difference.

4.2. Neural responses to attestedness and well-formedness

The ERP analysis revealed distinct neural responses according to the attestedness and the well-formedness of onset clusters. We found a larger negativity associated with ill-formed unattested clusters (compared to well-formed unattested clusters) during the 470–500 ms time window. We interpret this negativity as an N400 effect. The attested onsets (which were all well-formed) patterned with the well-formed unattested onsets at this time window, suggesting that the N400 effect was a response to phonologically ill-formed onsets (regardless of attestedness). At the time window around 600 ms, we found a greater positivity for unattested well-formed clusters compared to attested (well-formed) clusters. We interpret this positivity as an LPC effect. In this time later window, the unattested onsets (whether well-formed or ill-formed) patterned together to the exclusion of the attested onsets, suggesting that the LPC effect was a response to onsets that were unattested (regardless of well-formedness).

Following Ulbrich et al. (2016), we interpret the N400 effect observed in this study to represent pre-lexical phonological analysis of the stimuli (see also Deacon et al. 2004). The results suggest that upon hearing a potential word, participants first analyzed the onset cluster as a well-formed onset (i.e. an obstruent followed by an approximant in this case) or an ill-formed onset. Onsets recognized as ill-formed resulted in a greater N400 compared to those recognized as well-formed. We interpret the LPC effect to reflect the unfamiliarity of the unattested onsets following a lexical search. Previous studies have associated the LPC with the recognition of words and the familiarity of stimuli (e.g. van Petten & Senkfor 1996; Rugg & Curran 2007). Overall, these findings are consistent with a model in which potential words are first compared to a phonotactic grammar before being subjected to a full lexical search. Having phonological processing before lexical access could be functionally beneficial for speech perception. Recognizing words as highly improbable based solely on their phonotactics could allow the listener to avoid the processing costs associated with an exhaustive lexical search (cf. Luce & Pisoni 1998). Moreover, stepping back from isolated words to full phrases, this pre-lexical phonotactic analysis could facilitate word segmentation by inhibiting improbable parses thus making probable parses easier to find (e.g. see Moreton 2002).

Our results are largely consistent with the findings of Ulbrich et al. (2016), who also found distinct effects of well-formedness and attestedness in a similar study focused on coda clusters with German-speaking participants. As in this study, they found a greater N400

⁶ Note that Daland et al. (2011) also found better differentiation of the unattested clusters in their head-to-head comparison task than in their rating task.

effect for ill-formed clusters compared to well-formed clusters regardless of cluster attestedness, which they attributed to pre-lexical phonological processing. However, the LPC effect found in their study involved an interaction between attestedness and well-formedness, with clusters that were both unattested and ill-formed causing the largest LPC effect. In our study, there was no interaction of attestedness and well-formedness at this time window; ill-formed and well-formed unattested clusters patterned similarly to the exclusion of attested clusters. There were several differences between the studies that could have contributed to the difference in results. First, Ulbrich et al.'s study involved a word-picture matching task (conducted over multiple sessions) during the EEG recording. Our EEG recordings were taken during passive listening. The word-picture matching task required participants to actively learn the nonce words, to associate them with pictures, and to make a response each trial; it is possible that this task resulted in the interaction between well-formedness and attestedness in the later time window. Second, we tested onset clusters whereas Ulbrich et al. tested coda clusters. The different location of the target clusters within the words could have affected the nature of the processing in the task. With onset clusters, information about the well-formedness of the onset is available before the full identity of the word is available, but with coda clusters these bits of information become available simultaneously.

Another notable difference between Ulbrich et al.'s (2016) study and the current study is that almost all of the SSP-violating clusters used by Ulbrich et al. involved sibilant consonants, such as [s]. We chose not to include clusters containing sibilants in our ERP analysis because several researchers have theorized that these sequences are qualitatively distinct from typical consonant clusters due to the fact that they are cross-linguistically very common despite their poor sonority profile (Steriade 1982; van der Hulst 1984; Kaye 1992; Vaux 2004; Goad 2012).⁷ However, we did use a wider range of SSP-violating clusters as our unattested clusters, which allowed us to investigate the issue of gradience according to sonority profile.

We did not find evidence for a gradient neural response based on the SSP, even though our participants did exhibit a gradient sensitivity to the SSP in the behavioral task. We see several potential explanations for the lack of a gradient neural response. First, it is possible that the SSP is only relevant to processing that occurs earlier than the N400 time window. Perhaps any gradient distinctions motivated by the SSP are reduced to a more categorical well-formed or ill-formed status prior to the N400 time frame. This explanation would be consistent with the view that the graded SSP stems from the very early stages of perception rather than from phonological analysis *per se* (e.g. see Peperkamp 2007; Dupoux et al. 2011). We did not make hypotheses about effects earlier than N400 because there is little previous literature on effects related to phonotactics at earlier time windows, and earlier effects are more closely associated with perception and attention-driven effects than with phonological analysis. However, following a reviewer suggestion, we conducted a post-hoc exploratory analysis on a 170–200 ms time window for electrode CP6 and a 285–315 ms time window for electrode Fz, based on a visual inspection of the results (see Figure 3). We did not find any significant effects, but this possibility deserves attention in future studies.⁸

⁷ This was also why we were unable to include the fourth logically possible condition: attested clusters that violate the SSP. The only attested onsets in English that violate the SSP involve [s]. Ulbrich et al. (2016) were able to include this condition because they used such clusters.

⁸ For each time window, we used a one-way ANOVA comparing the three cluster groups (Attested Well-formed, Unattested Well-formed, and Unattested Ill-formed). There was no significant difference at 170–200 ms for electrode CP6 ($F(1, 9) = .766, p = .48$) or at 285–315 ms for electrode Fz ($F(1, 9) = .336, p = .72$).

A second possibility is that the gradient effect observed in the behavioral task stems from a later stage of processing, perhaps even as a type of meta-linguistic reanalysis. This explanation seems plausible for a nonce word rating task, but it would be unclear then why we see such gradient effects in relatively online perception tasks (e.g. Berent et al. 2007). Finally, it is possible that there is a gradient neural response, but it was too subtle to be detected in our experiment. In any case, more research is needed to better understand this seeming mismatch in the behavioral data and the ERP data.

A question that arises from this study is precisely which aspects of well-formedness the N400 seen in this study reflects. Sonority sequencing is not the only aspect of phonological well-formedness. For instance, many of our ‘well-formed’ clusters contained adjacent consonants with the same place of articulation (e.g. [tl], [pw]), another type of OCP violation, which is a possible explanation for why they are dispreferred in English and many other languages (e.g. Coetzee 2005). Domahs et al. (2009) found an ERP response to OCP-violating sCVC nonce words in German; however, the response was an LPC effect rather than an N400 effect. It is possible that some phonotactic violations, such as the restriction in English that onset clusters (other than those with sibilants) can only consist of obstruents followed by approximants, are processed at an earlier point than other types of violations, such as OCP violations. Such differences could also vary by language. Future work could explore this issue by looking at different types of violations within the same study.

Finally, it is worth considering how misperception could have influenced this study. Research has shown that listeners often perceptually repair sequences that are phonotactically illegal in their language (Dupoux et al. 1999; Berent et al. 2007; Davidson 2007; Peperkamp et al. 2008; Dupoux et al. 2011). If participants completely misperceived the stimuli in our experiment as CəCVC words rather than CCVC, then it raises concerns about whether the results accurately reflect the processing of the unattested clusters at all. This problem is a general point of concern with research of this type, sometimes leading researchers who conduct nonce word rating tasks to include orthography (alone or with audio) to aid participants in internalizing the intended form (e.g. Hayes & White 2013). This strategy comes with its own limitations, namely that it introduces orthotactic effects that might confound phonotactic effects. In the current study, we wanted to avoid orthotactic influences as much as possible by using an audio-only presentation while collecting the ERP data. We attempted to reduce the likelihood of misperceptions with our experimental design by telling participants that all words were a single syllable, and by including orthography in the initial familiarization phase and the pre-EEG rating task. We cannot be sure how participants perceived the stimuli during the passive listening task. However, the ratings in our post-EEG rating task (which did not include orthography) suggested that we are largely successful in helping the participants perceive the words as intended. Even if participants did misperceive some of the stimuli, the ERP results indicate that their neural responses were sensitive to cluster well-formedness at some level.

5. Acknowledgements

We would like to thank Andrew Clark for his assistance with the experiment, as well as two anonymous reviewers and the audience at OCP 13 for their comments. This project has been funded by a British Academy/Leverhulme Small Research Grant (round SRG 2015-16) to JW.

6. References

- Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26, 9–41.
- Ashburner, John et al. 2014. SPM12. The FIL Methods Group (and honorary members). URL <http://www.fil.ion.ucl.ac.uk/spm/>.
- Bailey, Todd M. and Ulrike Hahn. 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods. *Journal of Memory and Language* 44, 568–591.
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.
- Becker, Michael, Nihan Ketrez and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87, 84–125.
- Berent, Iris, Donca Steriade, Tracy Lennertz and Vered Vaknin. 2007. What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104, 591–630.
- Boersma, Paul and David Weenink. Praat: doing phonetics by computer [Computer program]. Version 6.0.22. URL <http://www.fon.hum.uva.nl/praat/>.
- Chomsky, Noam and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1, 97–138.
- Clements, G. N. 1992. The Sonority Cycle and syllable organization. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer and J. R. Rennison (eds.) *Phonological* 1988. Cambridge: Cambridge University Press. 63–76.
- Coetzee, Andries W. 2005. The OCP in the perception of English. In S. Frota, M. Vigario and M. J. Freitas (eds.) *Prosodies*. New York: Mouton de Gruyter. 223–245.
- Coleman, John and Janet B. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In J. Coleman (ed.) *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology*. Somerset, NJ: Association for Computational Linguistics. 49–56.
- Coulson, Seana. 2007. Electrifying results: ERP data and cognitive linguistics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey, (eds.) *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins. 400–423.
- Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28, 197–234.
- Davidson, Lisa. 2006. Phonology, phonetics, or frequency: influences on the production of non-native sequences. *Journal of Phonetics* 34, 104–137.
- Davidson, Lisa. 2007. The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24, 261–286.
- Deacon, Diana, Anna Dynowska, Walter Ritter and Jillian Grose-Fifer. Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology* 41, 60–74.
- Domahs, Ulrike, Wolfgang Kehrein, Johannes Knaus, Richard Wiese and Matthias Schlesewsky. 2009. Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech* 52, 415–435.
- Dupoux, Emmanuel, Kazuhiko Takehi, Yuki Hirose, Christophe Pallier and Jacques Mehler. 1999. Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568–1578.
- Dupoux, Emmanuel, Erika Parlato, Sonia Frota, Yuki Hirose and Sharon Peperkamp. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64, 199–210.
- Frisch, Stefan A. 1996. Similarity and frequency in phonology. Doctoral dissertation. Northwestern University.

- Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22, 179–228.
- Frisch, Stefan A. and Bushra Adnan Zawaydeh. 2001. The psychological reality of OCP-place in Arabic. *Language* 77, 91–106.
- Goad, Heather. 2012. sC clusters are (almost always) coda-initial. *The Linguistic Review* 29, 335–373.
- Goldsmith, John. 1979. *Autosegmental phonology*. New York: Garland.
- Hay, Jennifer, Janet B. Pierrehumbert and Mary E. Beckman. 2003. Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden and R. Temple (eds.) *Phonetic interpretation: papers in laboratory phonology VI*. Cambridge: Cambridge University Press. 58–74.
- Hayes, Bruce. 2011. Interpreting sonority-projection experiments: the role of phonotactic modeling. *Proceedings of the 2011 International Congress of Phonetic Sciences*. Hong Kong. 835–838.
- Hayes, Bruce and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44, 45–75.
- Hayes, Bruce and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.
- Hayes, Bruce, Kie Zuraw, Peter Siptár and Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85, 822–863.
- Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
- Huckvale, Mark. ProRec, version 2.3. URL <http://www.phon.ucl.ac.uk/resource/prorec/>.
- Hunter, Cynthia R. 2013. Early effects of neighborhood density and phonotactic probability of spoken words on event-related potentials. *Brain and Language* 127, 463–474.
- Hunter, Cynthia R. 2016. Is the time course of lexical activation and competition in spoken word recognition affected by adult aging? An event-related potential (ERP) study. *Neuropsychologia* 91, 451–464.
- Jarosz, Gaja and Amanda Rysling. 2017. Sonority sequencing in Polish: The combined roles of prior bias & experience. *Proceedings of the Annual Meeting on Phonology 2016*.
- Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig and Berlin: Teubner.
- Kaye, Jonathan. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics* 2, 293–313.
- Key, Alexandra P. Fonaryova, Guy O. Dove, and Mandy J. Maguire. 2005. Linking Brainwaves to the brain: An ERP primer. *Developmental Neuropsychology* 27, 183–215.
- Kutas, Marta and Kara D. Federmeier. 2011. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *The Annual Review of Psychology* 62, 621–647.
- Kutas, Marta and Steven A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207, 203–205.
- Kutas, Marta and Cyma K. Van Petten. 1994. Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (ed.) *Handbook of Psycholinguistics*. San Diego: Academic Press. 83–143.
- Luce, Paul A. and David B. Pisoni. 1998. Recognizing spoken words: The neighbourhood activation model. *Ear and Hearing* 19, 1–36.
- McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 51, 207–263.
- Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84, 55–71.

- Peperkamp, Sharon. 2007. Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition* 104, 631–637.
- Peperkamp, Sharon, Inga Vendelin and Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: experimental evidence from Japanese. *Phonology* 25, 129–164.
- Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In M. Aronoff and R. T. Oerhle (eds.) *Language sound structure*. Cambridge, Mass.: MIT Press. 107–136.
- Sievers, Eduard. 1881. *Grundzüge der Phonetik, zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel.
- Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Doctoral dissertation. MIT.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rossi, Sonja, Ina B. Jürgenson, Adriana Hanulíková, Silke Telkemeyer, Isabell Wartenburger and Hellmuth Obrig. 2011. Implicit processing of phonotactic cues: Evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience* 23, 1752–1764.
- Rugg, Michael D. and Tim Curran. 2007. Event-related potentials and recognition memory. *Trends in Cognitive Science* 11, 251–257.
- Ulbrich, Christiane, Phillip M. Alday, Johannes Knaus, Paula Orzechowska and Richard Wiese. 2016. The role of phonotactic principles in language processing. *Language, Cognition and Neuroscience* 31, 662–682.
- van der Hulst, Harry. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris.
- van Petten, Cyma and Ava J. Senkfor. 1996. Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. *Psychophysiology* 33, 491–506.
- Vaux, Bert. 2004. The appendix. Paper presented at Symposium on Phonological Theory: Representations and Architecture. CUNY, New York.
- Vitevitch, Michael S. and Paul A. Luce. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9, 325–329.
- Vitevitch, Michael S. and Paul A. Luce. 1999. Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory and Language* 40, 374–408.
- Vitevitch, Michael S. and Paul A. Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36, 481–487.
- Wiese, Richard, Paula Orzechowska, Phillip M. Alday and Christiane Ulbrich. 2017. Structural principles or frequency of use? An ERP experiment on the learnability of consonant clusters. *Frontiers in Psychology* 7.

6. Appendix

List of stimuli in orthography and IPA transcription.

Attested onsets:

<i>blace</i>	[bleis]	<i>frace</i>	[fɹeis]	<i>shrike</i>	[ʃɹaɪk]
<i>blike</i>	[blaɪk]	<i>frip</i>	[fɹɪp]	<i>shrip</i>	[ʃɹɪp]
<i>brip</i>	[bɹɪp]	<i>glace</i>	[gleis]	<i>smeak</i>	[smɪk]
<i>brome</i>	[bɹəʊm]	<i>glip</i>	[glɪp]	<i>smoon</i>	[smun]
<i>cleak</i>	[klik]	<i>grome</i>	[gɹəʊm]	<i>snace</i>	[sneis]
<i>clome</i>	[kləʊm]	<i>groom</i>	[gɹuːm]	<i>snike</i>	[snaɪk]
<i>crace</i>	[kɹeis]	<i>plome</i>	[pləʊm]	<i>swike</i>	[swaɪk]
<i>crike</i>	[kɹaɪk]	<i>ploon</i>	[plun]	<i>swip</i>	[swɪp]
<i>dreak</i>	[dɹɪk]	<i>prace</i>	[pɹeis]	<i>trome</i>	[tɹəʊm]
<i>drike</i>	[dɹaɪk]	<i>preak</i>	[pɹɪk]	<i>troon</i>	[tɹuːn]
<i>fleak</i>	[flɪk]	<i>queak</i>	[kwɪk]	<i>twip</i>	[twɪp]
<i>floon</i>	[flun]	<i>quome</i>	[kwəʊm]	<i>twoon</i>	[twun]

Marginal onsets:

<i>bwike</i>	[bwaɪk]	<i>shleak</i>	[ʃlɪk]	<i>thwace</i>	[θweɪs]
<i>bwip</i>	[bwɪp]	<i>shlike</i>	[ʃlaɪk]	<i>thwome</i>	[θwəʊm]
<i>dwreak</i>	[dwɪk]	<i>shmeak</i>	[ʃmɪk]	<i>vlip</i>	[vlɪp]
<i>dwoon</i>	[dwun]	<i>shmike</i>	[ʃmaɪk]	<i>vlome</i>	[vləʊm]
<i>fwace</i>	[fweɪs]	<i>shnace</i>	[ʃneɪs]	<i>vrip</i>	[vɹɪp]
<i>fwoon</i>	[fwun]	<i>shnome</i>	[ʃnəʊm]	<i>vrome</i>	[vɹəʊm]
<i>gwike</i>	[gwaɪk]	<i>shwace</i>	[ʃweɪs]	<i>vweak</i>	[vwɪk]
<i>gwoon</i>	[gwun]	<i>shwip</i>	[ʃwɪp]	<i>vwoon</i>	[vwun]

Unattested onsets:

<i>cmip</i>	[kmɪp]	<i>ltace</i>	[lteɪs]	<i>rdome</i>	[ɹdəʊm]
<i>cmoon</i>	[kmun]	<i>ltike</i>	[ltaɪk]	<i>rdoon</i>	[ɹduːn]
<i>dgace</i>	[dgeɪs]	<i>mlome</i>	[mləʊm]	<i>rgike</i>	[ɹgaɪk]
<i>dgike</i>	[dgaɪk]	<i>mloon</i>	[mlun]	<i>rgip</i>	[ɹgɪp]
<i>dnace</i>	[dneɪs]	<i>mrike</i>	[mɹaɪk]	<i>rleak</i>	[ɹlɪk]
<i>dneak</i>	[dnɪk]	<i>mrip</i>	[mɹɪp]	<i>rloon</i>	[ɹluːn]
<i>fnike</i>	[fnɪk]	<i>nleak</i>	[nɪk]	<i>rneak</i>	[ɹnɪk]
<i>fnome</i>	[fnəʊm]	<i>nlome</i>	[nləʊm]	<i>rnike</i>	[ɹnaɪk]
<i>lmeak</i>	[lmɪk]	<i>pkip</i>	[pkɪp]	<i>tlace</i>	[tleɪs]
<i>lmip</i>	[lmɪp]	<i>pkoon</i>	[pkun]	<i>tlome</i>	[tləʊm]
<i>lnace</i>	[lneɪs]	<i>pwace</i>	[pweɪs]	<i>zrip</i>	[zɹɪp]
<i>lnome</i>	[lnəʊm]	<i>pweak</i>	[pwɪk]	<i>zroon</i>	[zɹuːn]