

Title: Predictive coding of phonological rules in auditory cortex: A mismatch negativity study

Authors: Sari Ylinen^a, Milla Huuskonen^b, Katri Mikkola^b, Emma Saure^b, Tara Sinkkonen^b,
& Petri Paavilainen^{a,b}

^a Cognitive Brain Research Unit, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9, FIN-00014 Helsinki, Finland

^b Division of Cognitive and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9, FIN-00014 Helsinki, Finland

Corresponding author: Sari Ylinen, Cognitive Brain Research Unit, Institute of Behavioural Sciences, P.O. Box 9, FIN-00014 Helsinki, Finland;

e-mail: sari.ylinen@helsinki.fi

Abstract

The brain is constantly generating predictions of future sensory input to enable efficient adaptation. In the auditory domain, this applies also to the processing of speech. Here we aimed to determine whether the brain predicts the following segments of speech input on the basis of language-specific phonological rules that concern non-adjacent phonemes. Auditory event-related potentials (ERP) were recorded in a mismatch-negativity (MMN) paradigm, where the Finnish vowel harmony, determined by the first syllables of pseudowords, either constrained or did not constrain the phonological composition of pseudoword endings. The phonological rule of vowel harmony was expected to create predictions about phonologically legal pseudoword endings. Results showed that MMN responses were larger for phonologically illegal than legal pseudowords, and P3a was elicited only for illegal pseudowords. This supports the hypothesis that speech input is evaluated against context-dependent phonological predictions that facilitate speech processing.

Keywords: predictive coding, prediction error, phonological rules, vowel harmony, event-related potential (ERP), mismatch negativity (MMN)

1 Introduction

Predictive coding - the tendency of the brain to generate predictions of future sensory stimuli - represents a general principle of neural function that has been shown to manifest in auditory (Baldeweg, 2007; Wacongne et al., 2011), visual (Rao & Ballard, 1999; Summerfield, Trittschuh, Monti, Mesulam, & Eger, 2008), and sensory-motor (Hickok, Houde, & Rong, 2011; Ylinen et al., 2015) processing. Predictive-coding theory (Friston, 2009) proposes that sensory input is compared with predictions generated by a hierarchically organized predictive model to minimize surprise. In the hierarchical neural network, the predictive model is located higher in the hierarchy and sends its predictions to lower processing levels. Input matching the predictions will require less processing than mismatching input which generates a prediction error. The prediction error signal is projected to higher levels of the hierarchical network for updating the predictive model. The benefit of prediction is that there is no need to use full resources to process predicted input, whereas potentially important unpredicted events are processed further at higher levels (for discussion, see Bendixen, SanMiguel, and Schröger, 2012).

Interestingly, recent evidence suggests that predictive coding is also applied to speech processing: the brain continuously predicts future linguistic input based on the knowledge of one's native language. According to the magnetoencephalography (MEG) study by Gagnepain, Henson, and Davis (2012), future phonological segments are predicted on the basis of received speech input and known words. The authors trained their participants with novel words which were similar to familiar words, but with new endings past their prior uniqueness point. The training led participants to extend their expectations to include the sounds of novel words as they were added into their mental

lexicons. In the superior temporal gyrus (STG), this resulted both in an increased gradiometer field potential for the novel word and in a decreased field potential for the familiar word (Gagnepain et al., 2012), results not anticipated by prior lexical competition accounts (for these accounts, see Gaskell and Marslen-Wilson, 1997; McClelland and Elman, 1986; Norris and McQueen, 2008). In addition to predictions driven by lexical representations (see also Bendixen, Scharinger, Strauss, & Obleser, 2014), previous studies have suggested that predictions about following speech sounds may be generated on the basis of phonological knowledge (e.g., Hwang, Monahan, & Idsardi, 2010; Poeppel & Monahan, 2011; Scharinger, Idsardi, & Poe, 2011; Scharinger, Bendixen, Trujillo-Barreto, & Obleser, 2012; Weber, 2011).

In the auditory modality, predictive coding has been associated, among others, with the mismatch negativity (MMN; Näätänen, Gaillard, & Mäntysalo, 1978; for a review, see Näätänen, Paavilainen, Rinne, & Alho, 2007) component of the event-related potential (ERP; for predictive coding, see Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009; Wacongne et al., 2011; Wacongne, Changeux, & Dehaene, 2012; Bendixen et al., 2012). MMN is elicited by unexpected, rare deviant stimuli presented in the midst of a sound sequence otherwise obeying some regularity (the "oddball paradigm"). MMN is typically observed 150-250 ms after the onset of a deviant stimulus (Näätänen et al., 2007). The component originates from auditory cortex, and is elicited automatically even when attention is not directed to the auditory stimuli (Näätänen et al., 2007). According to the predictive coding interpretation of MMN (Friston, 2005; Winkler, 2007), the brain is continuously forming a model of the regularities of the auditory environment that creates predictions of future events. The MMN is elicited when the predictions are

violated and it is supposed to reflect a prediction error (Friston, 2005), updating the model (Winkler, 2007). Predictive coding account of MMN is also supported by a study by Wacongne et al. (2011) that used MMN to reveal hierarchical predictions of sound sequences in auditory cortex. In this study, predictive coding of non-speech sounds was shown to take place at multiple levels, creating hierarchical prediction errors in case of violation of two expectations.

Among other language-related phenomena (for reviews, see Näätänen, 2001; Pulvermüller & Shtyrov, 2006), MMN has been previously used to study phonological rules, including phonotactics (i.e., rules on the permissible phoneme combinations). For example, Dehaene-Lambertz, Dupoux, and Gout (2000) presented to French and Japanese listeners with sequences of pseudowords including phonological contrasts that were legal in French but illegal in Japanese. MMN was elicited for this contrast in French but not in Japanese listeners. According to the authors, the input signal is thus parsed into the phonological format of the native language. A similar phonological contrast was used by Jacquemot, Pallier, LeBihan, Dehaene, and Dupoux (2003) in a functional magnetic resonance imaging (fMRI) experiment, showing that phonological processing was associated with the activation of left superior temporal and left anterior supramarginal gyri. Further, MMN has been shown to reflect the assimilation rules of the place of articulation (Mitterer & Blomert, 2003; Mitterer, Csepe, Honbolygo, & Blomert, 2006; Tavabi, Elling, Dobel, Pantev, & Zwitserlood, 2009). More recently, Truckenbrodt, Steinberg, Jacobsen, and Jacobsen (2014) found no MMN for a consonant contrast concordant with the rule of final devoicing in German, whereas MMN was elicited for the same contrast when this rule was not applicable or when it was violated. In a similar

vein, Sun et al. (2015) studied MMN elicited by the voicing of voiceless consonants before certain, but not all, voiced consonants in French, suggesting sensitivity to complex phonological rules. Taken together, these studies show that phonological rules determine phonological parsing at the early stages of speech processing across languages and contrasts.

In spite of using MMN, the above mentioned studies leave open the question about the contribution of predictive coding to phonological processing under conditions where previous phonological units strongly constrain the selection of following units and thus may induce predictions about legal phonological units only. Although not explicitly discussed in the predictive coding framework, effects found by Steinberg, Truckenbrodt, and Jacobsen (2010a, 2010b, 2011) are relevant in this respect. The authors used MMN to explore the German phonotactic constraint of dorsal fricative assimilation with designs where predictions about following speech sounds may be induced: listeners were presented with phonotactically legal and illegal vowel-consonant combinations, where the vowel predicts legal consonants. Phonotactically ill-formed vowel-consonant deviants were found to elicit an enhanced or additional, later MMN response. This finding was attributed to the implicit phonotactic knowledge on which consonants can immediately follow certain vowels. This knowledge was interpreted to conflict with the auditory input, leading to violation detection and additional processing. More recent study by Steinberg, Jacobsen, and Jacobsen (in press) specified the effect of context on phonological repair and violation detection.

Phonological rules are abstract in nature, which means that they are applied on the basis of some phonological feature (e.g., the backness of vowels, the voicing of consonants). The above mentioned phonological MMN studies (e.g., Steinberg et al., 2010a, 2010b, 2011), however, share the feature of exploring adjacent phonemes that have different phonotactical co-occurrence probabilities. As a result, the possibility that MMN is affected by the co-occurrence probabilities of adjacent sounds (Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005) or co-articulatory cues (Steinberg, Truckenbrodt, & Jacobsen, 2012) is difficult to rule out entirely. A study design tapping the application of phonological rules on non-adjacent phonemes would help to tease apart acoustic-phonetic and abstract rule-based effects on MMN, because non-adjacent phonemes should be less prone to the effects of co-occurrence probabilities and co-articulation. This kind of data could thus provide further support for the abstract nature of previously observed phonotactic MMN effects. In addition, the study of P3a (or novelty-P3) response, reflecting involuntary attention shift to the stimulus deviance in MMN paradigms (Escera & Corral, 2007; Polich, 2007) could further illuminate the processing of phonological constraints. To this end, we measured ERPs and specifically MMN and P3a responses with the aim to determine whether predictive coding is applied to the processing of language-specific phonological rules that constrain non-adjacent phoneme sequences in word forms (i.e., the phonological forms of words or pseudowords).

The specific phonological rule chosen for the present study is the remarkably consistent vowel harmony of the Finnish language (for a review, see Karlsson, 1983). According to the rules of the vowel harmony, front vowels (/æ/, /ø/ and /y/) and back vowels (/ɑ/, /o/

and /u/), named by the different positions of the tongue during articulation, may never occur in the same word, whereas all vowels can occur with neutral vowels /e/ and /i/. Vowel harmony affects also word inflection by determining the choice of allomorphic inflectional affixes (e.g., /talo+ssa/ 'in a house' vs. /møki+ssæ/ 'in a hut'). The rule is very prevalent in Finnish, and therefore it is used as a cue to segment words from continuous speech. Practically the only exceptions to this rule are compound words and loan words from foreign languages. As a result, many Finns find it difficult to correctly pronounce foreign loan words violating the vowel harmony, such as [olympia] 'Olympic'.

The Finnish vowel harmony has been previously studied with MMN by Aaltonen et al. (2008; see also Scharinger, Idsardi, & Poe, 2011, for vowel harmony in Turkish). The authors compared MMNs between two groups, namely, the native speakers of Finnish and Estonian. Estonian belongs to the same Uralic language family as Finnish but lacks the vowel harmony. The Finnish and Estonian speakers were tested with a standard stimulus [tækæ] and a deviant stimulus violating the Finnish vowel harmony. The deviant stimulus featured at the end of [tæk] a non-native vowel, an intermediate between the Finnish /a/ and /æ/ vowels. The deviant stimulus elicited an enhanced MMN response in Finns compared to Estonians, which was interpreted as reflecting the detection of native-language rule violations in Finns. However, the response reported as MMN peaked 300 ms after the onset of the critical vowel. This would be an unusually long latency for MMN (Näätänen et al., 2007), complicating the interpretation of the results. Moreover, the study used a non-prototypical vowel as the deviant stimulus which could possibly result in differences in the responses between the two groups (see Näätänen et al., 1997). This drawback could have been eliminated by demonstrating a

significant interaction between the critical pseudoword condition and a control condition of isolated vowels, but no such interaction was found, failing to rule out alternative interpretations of the data. Potentially, results showing an MMN enhancement in the very same participants for a vowel violating the vowel harmony but no enhancement for the same vowel in a phonologically legal context would provide stronger evidence for predictive coding of phonotactic/phonological rules than results comparing different language groups.

In the present study, native Finnish speakers were presented with spoken pseudowords or non-words in a passive oddball paradigm where the participants ignored the auditory stimuli. The initial vowel of the pseudowords either did or did not constrain the non-adjacent final vowel by the rule of the Finnish vowel harmony. To probe the predictions derived from the initial vowel, we used word forms with deviant final vowels that were either consistent with the Finnish vowel harmony or violated it. Thus, in the *neutral condition*, the stimuli started with the vowel-harmony neutral vowel [i] ("the /i/ context"). The standard word form was [ika] and the deviants were [iko] and [ikø] (both phonologically legal in Finnish). In the *vowel harmony condition*, the stimuli started with the back vowel [ɑ] ("the /ɑ/ context"). The standard was [ɑka] and the two deviant stimuli were [ɑko] with back vowels, obeying the vowel harmony, and [ɑkø]* with back and front vowels, violating the vowel harmony (this phonologically ill-formed stimulus will be marked with an asterisk from here on). Consequently, the present design enabled to compare MMNs to the same vowel [ø] in phonologically legal and illegal contexts.

Our hypotheses are based on observations that predictability can be induced by various sources, such as the history of stimulation and long-term memory (see Bendixen et al., 2012, for a review). In the framework of hierarchical predictive coding (see Friston, 2009; Wacongne et al., 2011) applied to the processing of speech (Ylinen et al., submitted), the regular presentation of a word form in an auditory stimulus sequence was hypothesized to lead to the predictions of hearing the same word form. Thus, the deviation of the final vowel was expected to elicit a sequence-level MMN (in terms of Bendixen et al., 2012, the extraction of regularities is here based on auditory–auditory links; see Poeppel and Monahan, 2011, for discussion on local predictions). However, as commanded by the rule of vowel harmony, a word-form initial back vowel was hypothesized to create the predictions of legal vowels (i.e., vowels from the same or neutral backness category) in the word-form final position. This would constitute another level of hierarchical predictions that is based on abstract phonological rules and would concern non-adjacent phonemes. The violation of the vowel harmony rule, established in long-term memory, was expected to result in a phonological prediction error adding up to the MMN for sequence-level predictions. In contrast, the presence of a neutral vowel at the beginning of a word form was hypothesized to lead to sequence-level predictions, but not to specific phonological predictions, as neutral vowels do not generate phonological constraints. Thus, we hypothesized that the MMN elicited by word forms violating the vowel harmony rule is larger than the MMN for phonologically legal deviants (in line with Steinberg et al., 2010a). In addition, we hypothesized that the violation of a strongly established phonological rule may result in a switch of attention, which is reflected by the P3a response.

2 Methods

2.1 Ethics statement

The study was approved by the University of Helsinki Ethical review board in humanities and social and behavioral sciences. All participants gave their informed written consent.

2.2 Participants

Fifteen participants (8 women, 7 men; aged 18-29, average 24.9, SD 2.87) volunteered in the experiment. All were right-handed and reported no neurological or psychiatric disorders or language impairments. All participants spoke Finnish as their native language and had grown up in an exclusively Finnish-speaking household.

2.3 Stimuli

The stimuli were designed as pseudowords or non-words in order to avoid any semantic MMN effects (Pulvermüller et al., 2001). A female native speaker of Finnish pronounced phonologically legal pseudowords /aka/, /ako/, /ækø/, /ika/, /iko/ and /ikø/ several times in a sound-attenuated chamber, while these utterances were recorded with high-quality sound-processing equipment [Mac OS X 10.7.1 computer (Apple, Cupertino, CA), Pro Tools 10 software (version 10.0.0; Avid Technology Inc., Twinsburg, OH), Digidesign Digi 002 digital audio workstation (Avid Technology Inc., Twinsburg, OH), and AKG C 4000B microphone (AKG Acoustics GmbH, Vienna)]. The first syllables and the second syllables (from the release burst of /k/ onwards) were isolated from these natural recordings and presented to three native speakers, naïve to the origins of the stimuli. They could not correctly guess the context from which the final syllables had been isolated. Based on their judgements, most prototypical syllables ([ɑ] and [i] for

pseudoword onsets and [kɑ], [ko] and [kø] for their endings) were selected. The onset syllables were chosen from pseudowords [ɑkɑ] and [ikɑ] to avoid any possible bias with respect to the other two endings used as deviants. The syllables [kɑ], [ko] and [kø] originated from pseudowords /ika/, /iko/ and /ikø/. As measured in 50 ms time windows immediately following vowel onset, the formant frequencies F1-F3 of these pseudoword-final vowels were 677 Hz, 1280 Hz, and 2408 Hz for [ɑ], 537 Hz, 1026 Hz, and 2379 Hz for [o], and 461 Hz, 1824 Hz, and 2309 Hz for [ø], respectively (see Wiik, 1965, for similar formant values). It is noteworthy that according to the formant values, the acoustical difference between the standards [ɑkɑ]/[ikɑ] and the deviants [ɑkø]/[ikø] was larger than that between the standards and the deviants [ɑko]/[iko]. Since this inevitably affects the MMN amplitude, our main focus of interest was to compare the responses to the identical final syllables.

After selection, the initial and final syllables were combined into two triplets of bisyllabic pseudoword or non-word stimuli: [ikɑ], [iko], [ikø] and [ɑkɑ], [ɑko], [ɑkø]* (see Fig. 1). In this set of stimuli, each initial vowel was combined with the three final syllables. The initial vowel was thus identical within each triplet. In addition, each final syllable was identical across the triplets used in different conditions (e.g., the final syllable was identical in [ikø] and [ɑkø]*). The duration of the silent occlusion phase of /k/, serving as the point of cross-splicing, was kept similar to that in original recordings. The syllables with the same position (initial and final) were edited in Adobe Audition 2.0 (Adobe Systems Inc., San Jose, CA) and in Praat (Boersma & Weenink, 2010) to have similar durations, fundamental frequency (F0) and intensity, preserving the natural ratios between the syllables and the silent occlusion phase. As measured 50 ms following the

vowel onset, the intensity and F0 of the onset vowels were 83 dB and 197 Hz, respectively, and those for the final vowels [ka], [ko], [kø] were 85, 85 and 84 dB and 187, 184 and 183 Hz, respectively. After editing, the stimuli were resynthesized with the overlap-add function of Praat (Boersma & Weenink, 2008). The beginnings and endings of each stimulus were attenuated with 5 ms ramps. The intensity of the occlusion phase of the stop consonant /k/ was set to zero between the initial vowel and the release burst of the consonant using 5 ms ramps in both ends of the occlusion phase.

The total duration of the stimuli was 380 ms. The duration of the initial vowel was 100 ms, that of the occlusion phase of the stop consonant /k/ 80 ms, and that of the release burst and the final vowel 200 ms. Because in all stimuli the release burst of the consonant was aligned to the time point of 180 ms and there was some variance in voice onset time (VOT), the exact onset of the final vowel slightly differed between the stimuli. For [ika], [iko], [aka], and [ako], the duration of the release burst was 20 ms and that of the final vowel 180 ms, whereas for [ikø] and [akø]* the duration of the release burst was 14 ms and that of the final vowel 186 ms (the 6 ms difference in VOT could, in principle, affect the MMN, but that is controlled for by comparing the responses between [ikø] and [akø]* with identical timing). The standard and the deviant stimuli physically differed from each other from 180 ms onwards (i.e., from the release burst onwards). However, since the consonant was /k/ in all cases, it is not excluded that the deviation could be detected after the vowel onset, which took place around 200 ms (at 194 ms for [ikø] and [akø]* and at 200 ms for the other stimuli).

Since speech processing is modulated by co-articulation (Steinberg et al., 2012) and the chosen final syllables [kɑ], [kɔ] and [kø] had been produced in the context of /i/, a question rises whether the initial /i/ vowel could have generated co-articulatory cues that carry over the consonant to the final vowel. To examine possible co-articulatory effects in our original natural recordings, we measured the formant frequencies of [ɑkɑ], [ɑkɔ], [ækø], [ikɑ], [ikɔ], [ikø] (6 exemplars from each) in 50 ms time windows immediately following the final vowel onset. The formant values of the [ɑ] and [ɔ] vowels of our experimental stimuli, taken from /i/ context, fell within the range of F1 and F2 formant values of the vowels pronounced in the /a/ context (F1-F3 ranges for the final [ɑ] of [ɑkɑ]: 670-749 Hz, 1193-1292 Hz, 2238-2366 Hz; F1-F3 ranges for [ɔ] of [ɑkɔ]: 522-567 Hz, 965-1027 Hz, 2280-2407 Hz). This was not the case with [ø] taken from [ikø], compared against formant values of [kø] following [æ] (F1-F3 ranges for [ø] of [ækø]: 492-564 Hz, 1544-1730 Hz, 2148-2292 Hz; note that no illegal /ɑkø/* was recorded at all). However, the differences between the vowel of the chosen [kø] syllable and the [ø] vowels from [æ] context were so small (F1: 31 Hz, F2: 93 Hz, F3: 17 Hz) that they were below just noticeable difference (JND; according to Mermelstein and Fitch, 1976, JNDs for steady state vowels are 50 Hz for F1 and 142 Hz for F2 and JNDs for vowels in consonantal context are 46-51 Hz for F1 and 174-199 Hz for F2; no JNDs for F3 were reported). Using the same speech samples, we also examined the effect of preceding vowel on the acoustic characteristics of the stop consonant release bursts by measuring the center of gravity of their spectra in /i/ and /a/ contexts. There was substantial overlap in the spectral quality of /k/ in each context: 7013 vs 7797 Hz for [ɑkɑ] vs [ikɑ], 6749 vs 6670 Hz for [ɑkɔ] vs [ikɔ], and 8037 vs 8069 Hz for [ækø] vs [ikø]. Thus, the acoustical

measurements were in line with the judgements of our 3 naïve listeners who could not identify or guess the context of the syllables on the basis of co-articulation.

2.4 Experimental design and ERP recording

The experiment had two conditions. In the *vowel harmony condition*, the standard stimulus was the pseudoword [akɑ] ($p=0.80$). The deviant stimulus consistent with the vowel harmony was [ako] ($p=0.10$) and the deviant stimulus violating the vowel harmony was [akø]* ($p=0.10$). In the *neutral condition*, where the initial vowel did not constrain the following vowels, the standard stimulus was [ika] ($p=0.80$). The deviant stimuli were [iko] ($p=0.10$) and [ikø] ($p=0.10$), which were concordant with the vowel harmony. The only differences between stimuli in the /ɑ/ and /i/ contexts were the vowels at the beginning of the pseudowords, and thus the responses to stimuli with the same final syllable ([akø]* vs. [ikø]) could be directly compared between the contexts to isolate the effect of vowel harmony, whereas the comparison between [ako] and [iko] served as a control for the effect of the initial vowel.

The experiment consisted of six stimulus blocks (three random sequences per each condition). The order of the blocks was counterbalanced for each participant and across participants. Each sequence contained 400 standard stimuli and 50 deviant stimuli of each type and lasted about 7 minutes. The stimulus-onset asynchrony (SOA) was 900 ms. The stimuli were presented in a pseudorandom order, with the restriction that there were at least two standard stimuli between subsequent deviant stimuli. The responses to the first 4 stimuli of each sequence and the two standard stimuli following each deviant were rejected from averaging.

In some MMN studies, participants are presented with stimulus blocks with reversed probabilities of standard and deviant stimuli to be able to generate difference waveforms where responses to identical stimuli have been subtracted from each other (see, e.g., Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006; Steinberg et al., 2010a). This is to minimize the unequal contribution of exogenous activation (e.g., N1) for different stimuli in the difference waveforms. We considered not necessary to run such stimulus blocks, because the exogenous activation for syllables ending with /o/ and /ø/ was expected to be the same across conditions. Therefore, a significant interaction between condition and final syllable, which we expected to find, could not be caused by different exogenous activation, regardless of using responses to different vowels in the subtraction.

EEG was recorded with Ag/AgCl electrodes from Fz, Cz, Pz, F3, F4, C3, C4, LM (left mastoid) and RM (right mastoid). Vertical eye movements (VEOG) were recorded with two electrodes placed above and below the left eye. Horizontal eye movements (HEOG) were recorded with two electrodes placed at the outer canthi of the left and right eyes. The reference electrode was placed on the tip of the nose and the grounding electrode on the forehead. The EEG was recorded with a sampling rate of 200 Hz using an online band-pass filter of 0.10-40 Hz. The stimuli were presented binaurally using headphones. The participants were instructed to concentrate on watching a self-selected silent movie with Finnish subtitles while ignoring the sounds from the headphones. Between the blocks, the participants could take a short break whenever they wished. Including the breaks, the experiment proper lasted about 50 minutes per participant.

2.5 Data analysis

The EEG was filtered (pass band 1-30 Hz, roll-off 24 dB/octave), epoched using the analysis period of -100-600 ms, and ERPs were averaged separately for all stimulus types. Trials including eye blinks or other artifacts exceeding $\pm 100 \mu\text{V}$ were excluded from the averaging. The prestimulus period of -100-0 ms was used as a baseline for amplitude measurements. Difference waveforms were calculated by subtracting the ERPs to standard stimuli from those to deviant stimuli.

The MMN and P3a mean amplitudes were measured from the difference waveforms in 40 ms time windows centred around peak latencies at Cz where the maximal or almost maximal amplitudes were observed for P3a and MMN, respectively. Peak latencies were determined separately for each deviant type to ensure that the contribution of maximal amplitudes was equal across stimulus types. The MMN peaks were observed at 340 ms for [ako], 350 ms for [akø]*, 345 ms for [iko], and 365 ms for [ikø].

Correspondingly, the time windows of mean amplitude measurement were 320-360 ms for [ako], 330-370 ms for [akø]*, 325-365 ms for [iko], and 345-385 ms for [ikø]. The P3a peaks (or the most positive deflections following the MMN) were observed at 445 ms for [ako], 440 ms for [akø]*, and 435 ms for [iko] and [ikø]. Correspondingly, the time windows of mean amplitude measurement were 425-465 ms for [ako], 420-460 ms for [akø]*, and 415-455 ms for [iko] and [ikø]. To test the significance of MMN and P3a responses, mean amplitudes averaged across fronto-central scalp sites (including F3, Fz, F4, C3, Cz, and C4) were compared with zero in t-tests (one-tailed, since the polarity of MMN and P3a responses was known in advance). In a similar vein, t-tests

were used to test the significance of MMN polarity reversal, averaged across mastoid sites (including left and right mastoids). The mean amplitudes for MMN and P3a were analyzed with separate 4-way repeated-measures ANOVAs with factors Condition (vowel harmony vs. neutral), Final syllable ([ko] vs. [kø]), Coronal scalp site (frontal vs. central), and Sagittal scalp site (left vs. midline vs. right). Significant interactions were further investigated with a priori defined pairwise comparisons (ako vs. iko; akø vs. ikø). Bonferroni correction for multiple comparisons was applied to pairwise comparisons following interactions and main effects with more than two factors. Effect sizes were determined using partial eta squared and Cohen's d. In all comparisons, alpha level was 0.05. Only those interactions involving factors Condition and Final syllable are reported, since the other interactions are not related to our research questions.

3 Results

ERPs to the standard and deviant stimuli in the vowel harmony and neutral conditions are shown in Figure 2 and MMN mean amplitude values at fronto-central sites (F3, Fz, F4, C3, Cz, and C4) and mastoids in Table 1. The deviant stimuli elicited MMNs, differing significantly from zero and peaking at about 350 ms (i.e., about 170 ms after the physical-change onset and about 150 ms after the final-vowel onset). Also responses at mastoids, showing reversed polarity, differed significantly from zero.

Table 1. Mean MMN amplitudes (\pm standard deviations), averaged across fronto-central (FC) scalp sites, including F3, Fz, F4, C3, Cz, and C4, and across mastoid (M) sites, including left and right mastoids.

	MMN mean amplitude (μ V), FC	t value (df 14), FC	p value, FC	MMN mean amplitude (μ V), M	t value (df 14), M	p value, M
ak \emptyset *	-2.58 (\pm 1.78)	-5.62	<0.001	0.92 (\pm 0.99)	3.62	<0.0025
ako	-0.78 (\pm 0.86)	-3.52	<0.005	0.40 (\pm 0.54)	2.86	<0.01
ik \emptyset	-1.48 (\pm 1.34)	-4.26	<0.001	0.74 (\pm 0.44)	6.59	<0.001
iko	-0.53 (\pm 0.83)	-2.49	<0.025	0.22 (\pm 0.43)	2.01	<0.05

Table 2. Mean P3a amplitudes (\pm standard deviations), averaged across fronto-central (FC) scalp sites, including F3, Fz, F4, C3, Cz, and C4. n.s. refers to not significant.

	P3a mean amplitude (μ V), FC	t value (df 14), FC	p value, FC
ak \emptyset *	1.69 (\pm 1.62)	4.03	<0.001
ako	0.14 (\pm 1.28)	0.44	n.s.
ik \emptyset	0.19 (\pm 1.11)	0.68	n.s.
iko	0.33 (\pm 1.00)	1.27	n.s.

Figure 3 shows the difference waveforms for the same final syllables ([kø] and [ko]) superimposed in different phonological contexts. An ANOVA for MMN mean amplitude revealed a significant interaction of Condition x Final syllable [$F(1,14) = 4.584, p < 0.05, \eta_p^2=0.247$]. Pairwise comparisons following up the interaction showed that the MMN was significantly larger for [akø]*, violating the vowel harmony, than for [ikø] with the same final syllable in a neutral context [$t(14)=3.577, p=0.003, d=0.92$]. In contrast, no significant difference was found between the MMN responses to [ako] and [iko] [$t(14)=1.259, p=n.s., d=0.32$]. In addition to the interaction, all main effects were significant: Condition $F(1,14)=16.16, p=0.001, \eta_p^2=0.536$; Final syllable $F(1,14)=28.82, p<0.001, \eta_p^2=0.673$; Coronal scalp site $F(1,14)=8.21, p=0.012, \eta_p^2=0.370$; Sagittal scalp site $F(1,14)=5.39, p=0.01, \eta_p^2=0.278$. The effects of Condition and Final syllable were in more detail explained by the Condition x Final syllable interaction. The effect of Coronal scalp site was due to larger responses at central than frontal scalp sites, whereas that of Sagittal scalp site was due to larger responses at midline compared to the right scalp sites [$t(13)=3.036, p=0.027, d=0.78$].

MMN to [akø]* was followed by a positive deflection peaking at about 400 ms, interpreted as P3a on the basis of its timing, polarity, and maximal amplitude at central scalp sites. Unlike MMN which was observed for all stimulus types, P3a was observed only for [akø]*: the mean amplitude for [akø]* differed significantly from zero at fronto-central sites, whereas P3a was not significant for any other stimulus type (see Table 2 for mean amplitude values). This was substantiated by a significant ANOVA interaction of Condition x Final syllable [$F(1,14)=25.85, p<0.001, \eta_p^2=0.649$]. Follow-up

comparisons indicated that the deviant [akø]* elicited a significantly larger positivity than [ikø] [$t(14)=5.28$, $p=0.005$, $d=0.847$], whereas no significant difference was found between the other contrast [ako] and [iko] [$t(14)=0.611$, $p=n.s.$, $d=0.157$]. In a similar vein, Condition x Final syllable x Coronal scalp site [$F(1,14)=29.47$, $p<0.001$, $\eta_p^2=0.678$] and Condition x Final syllable x Sagittal scalp site [$F(1,14)=6.49$, $p=0.005$, $\eta_p^2=0.317$] interactions were significant. The follow-up comparisons of these interactions again indicated a significantly larger positivity for [akø]* than [ikø] [for all comparisons, $t(14)>2.47$, $p<0.05$, $d>0.639$], whereas none of the comparisons were significant for the other contrast [ako] and [iko] [$t(14)<1.36$, $p=n.s.$, $d<0.351$]. In addition, ANOVA indicated significant main effects of Final syllable [$F(1,14)=10.21$, $p=0.006$, $\eta_p^2=0.422$], Coronal scalp site [$F(1,14) = 7.41$, $p=0.017$, $\eta_p^2=0.346$], and Sagittal scalp site [$F(1,14) = 3.50$, $p=0.044$, $\eta_p^2=0.20$]. The effect of Final syllable was in more detail explained by the interactions. The effect of Coronal scalp site was again due to larger responses at central than frontal scalp sites. None of the comparisons involving Sagittal scalp site was significant after Bonferroni correction [$t(13)<2.17$, $p=n.s.$, $d<0.560$].

4 Discussion

By comparing ERP difference responses, the present study aimed at determining whether hierarchical predictive coding is applied to the processing of language-specific phonological rules that constrain phoneme sequences comprising potential words. The main finding was that as suggested by the significant interactions between the context and the final syllable, the MMN and P3a amplitudes for the final-syllable vowel were determined by the previous vowel only when the final vowel violated the vowel harmony rule. The interactions were driven by enhanced responses to the illegal [akø]*.

Since acoustical differences between deviant and standard stimuli exert influence on MMN amplitude, it is reasonable to address the possibility that simple acoustical effects could explain the current pattern of results. MMN was larger for [kø] than [ko] syllables, which is likely due to larger acoustical discrepancy from the standard [ka] (for formant frequencies, see Section 2.3). However, we were mostly interested in comparing the responses to the same syllables in different contexts (vowel harmony vs. neutral). Acoustic features or any non-contextual saliency of the critical final syllable do not account for the MMN enhancement for [akø]*, because no such effect was observed for the identical syllable in the neutral condition ([ikø]), where the initial vowel did not constrain the final vowel. Neither does potentially different exogenous activation caused by the initial vowel¹ explain the significantly larger MMN for [akø]* than for [ikø], since no significant difference was found between the MMNs for [ako] and [iko]. Potentially, MMN enhancement could also be related to the phonotactical co-occurrence probability of phonemes, because higher phonotactical co-occurrence probability of adjacent sounds has been suggested to result in larger MMN as compared with lower phonotactical co-occurrence probability (Bonte et al., 2005). However, such co-occurrence probabilities cannot explain the current pattern of results, where MMN enhancement was found for the illegal phonotactic sequence with zero probability (in contrast, phoneme triplets /ako/, /iko/ and /ikø/ occur in Finnish words, e.g., /pako/ 'escape', /liko/ 'soak', /eikø/ 'no?')

¹As illustrated by Fig. 2, the N1 for the first syllable [a] (at 120 ms) is smaller, that is, more positive than the N1 for the first syllable [i]. This is likely due repetition suppression, since neural populations responsive to /a/ need to be more frequently active when the standard stimulus is [aka] than when it is [ika]. However, this effect ceases at 300 ms, that is, before the MMN time window. At MMN latency (320 ms onwards), the response to the final [a] of [aka] is, if anything, more negative than that to the final [a] of [ika]. Thus, the subtraction of the response to standard [aka] from the response to deviant [akø]* could not artificially cause the MMN enhancement.

or 'not?'). Thus, the processing of phonologically illegal and legal but infrequent sound combinations seems to be fundamentally different (Steinberg et al., in press).

Although acoustic features, exogenous activation, or phonotactical co-occurrence probabilities cannot account for the present results, the findings are compatible with the hypothesis of hierarchical predictive coding (see Friston, 2009; Wacongne et al., 2011). We expected that on the basis of vowel harmony rule, pseudoword-initial vowels would enable predicting the following vowels at the word-form level of the hierarchy, independently of regularities in the stimulus sequence (Poeppe & Monahan, 2011, call these knowledge-based predictions, see also Ylinen et al., submitted). The predictions were expected to affect the processing of the final syllables, which is compatible with the present data, as substantiated by the significant interaction between the context and the final syllable. In line with the predictive coding hypothesis, the unpredicted illegal word form elicited an enhanced MMN, whereas the MMN responses to predicted phonotactically legal word forms were of smaller amplitude. This MMN enhancement is interpreted as a prediction error signal at the level of word forms (see Wacongne et al., 2012, for a detailed neurophysiological account on prediction-error MMN in a non-linguistic context).

Since in the present study we used pseudo- and non-words rather than words, our effects were not lexical. As discussed above, neither is the data pattern explained by acoustic or phonetic processing. Thus, in line with previous studies suggesting that predictions about following speech sounds may be generated on the basis of phonological knowledge (e.g., Hwang et al., 2010; Poeppe & Monahan, 2011;

Scharinger et al., 2011, 2012; Weber, 2011), we can conclude that our findings were driven by phonology. This interpretation is compatible with the observation of maximal amplitudes at central electrodes, since also Steinberg et al.'s (in press) recent study on the effect of context on phonologically driven MMN focused on central scalp sites. The fact that inverted polarity was observed at the mastoid electrodes suggests that the response at least partly originated from auditory cortex, known to be the major source of MMN (Näätänen, 2001; Näätänen et al., 2007). Thus, our pattern of findings seems best explained by a neural network that involves auditory cortex, that is sensitive to phonological context (Steinberg et al., in press), and that predicts phonologically or phonotactically legal items on the basis of language-specific phonological rules applied from long-term memory (e.g., Dehaene-Lambertz et al., 2000; Steinberg et al., 2010a).

In a hierarchical predictive coding framework, the MMN is certainly not only contributed by the word-form level that is proposed to be independent of regularities in the stimulus sequence, but also sequence-level predictions (in terms of Monahan & Poeppel, 2011, local predictions). These refer to predictions about hearing the most frequent items of the stimulus sequence (the standards). Although word-form processing was proposed to be determined by long-term memory, the MMN amplitude is most likely modulated by long-term memory at the sequence level as well, since familiar or prototypical items may have stronger brain representations compared to unfamiliar or atypical items (e.g., see Näätänen et al., 1997; Bonte et al., 2005; Ylinen et al., 2006). It is also well-established that at this level of processing the MMN is affected by the acoustical discrepancy between the stimuli. In the present study, smaller MMNs for phonotactically legal stimuli are proposed to be driven mainly by the sequence-level processing and to reflect

suppressed word-form-level responses. Small differences in the response amplitude for these legal stimuli are likely induced by acoustical differences between the standards and the deviants (see Table 1, [iko] vs. [ikø] for MMN amplitudes and Section 2.3 for acoustical differences between the stimuli). A very small difference (0.25 μ V) was also observed between the responses to [ako] and [iko], which could be due to vowel context at sequence level, but this effect was non-significant. Thus, although sequence-level processing inevitably affects MMN amplitudes, in the present study its contribution to the overall pattern of findings is non-significant and thus marginal (i.e., it does not account for the interaction between the context and the final syllable).

Previous studies on MMN and phonological rules (e.g., Dehaene-Lambertz et al., 2000; Steinberg et al., 2010a, 2010b, 2011) have interpreted their phonological effects on MMN as reflecting implicit phonological or phonotactical knowledge. This interpretation is in agreement with the current findings, since long-term memory representations for phonological rules learned implicitly as a result of exposure to native language are certainly required for accurate predictions about the phonological structure. The framework of hierarchical predictive coding may explain how this phonological knowledge could be applied to the processing of speech input. Specifically, as proposed above, the comparison of speech input with rule-based predictions facilitates processing, because the comparison of the input with a certain prediction is easier than the recognition of random unpredicted input. Furthermore, only input mismatching the predictions and eliciting prediction error requires additional processing resources. The constraints set by phonological rules may thus function to enable faster and more efficient processing of auditory input, similarly as existing long-term memory

representations for words facilitate the processing of spoken words (Gagnepain et al. 2012; Ylinen et al., submitted).

Importantly, our results extend previous findings by demonstrating the effect of phonological rules on the processing of non-adjacent speech segments. Compared with adjacent segments, non-adjacent segments are less prone to the effects of phonotactical co-occurrence probabilities and less affected by co-articulation (see Section 2.3). This emphasizes that speech processing is modulated by phonological rules that are most likely abstract in nature. The stimulus-initial back vowel in the vowel harmony condition may have resulted in the prediction of encountering any legal successor (any back or neutral vowel) or the prediction of not encountering any illegal successor (in the vowel harmony condition, any front vowel). These alternatives may represent the two sides of the same coin, because the pre-activation of some items may inhibit others. In contrast with the vowel harmony condition, any final vowel is allowed in the neutral condition, and therefore no specific predictions about vowels may have occurred.

In addition to MMN, phonologically illegal word forms, but not the other deviants, elicited a positive ERP response, which was interpreted as a P3a because of its latency and fronto-central scalp distribution (Polich, 2007). P3a is thought to indicate an involuntary attention shift to the stimulus deviance and it often follows strong MMN responses (Escera & Corral, 2007; Polich, 2007). The elicitation of P3a only by the stimulus that

violated the Finnish vowel harmony suggests that the deviation from phonological expectations was not only automatically detected, but it also led to an involuntary attention shift to the unpredicted word form. This attention shift may have been triggered by the prediction error (Schröger, Marzecová & SanMiguel, 2015; Ylinen et al., submitted). According to the predictive coding hypothesis, the function of the prediction error is to inform higher levels of processing about the violation of predictions (Friston, 2009). Applied to the present data, the violation of the phonological rule could result in the allocation of additional resources to process the unexpected stimulus. Consequently, the present phonologically illegal deviants led not only to quantitative, but also to qualitative processing differences.

Recently, Chennu et al. (2013) explored the effects of predictability and attention on MMN and P300 (i.e., P3b) responses to non-speech sounds. They found that the MMN is sensitive to stimulus predictability and diminished by top-down expectations, whereas the P300 is sensitive to attentional engagement and sharpened by top-down expectations, suggesting different levels of predictive complexity. Although Chennu et al.'s (2013) experimental design with attentional manipulations was essentially different than ours, the current MMN and P3a results are in accordance with their findings on different levels of processing complexity, with MMN reflecting more fine-grained processing and P3a functioning on a coarser-grain (all-or-nothing) manner. The MMN and P3a results of the present study are also compatible with previous findings suggesting that large MMN and P3 responses indicate rule learning from auditory input (Mueller, Friederici, & Männel, 2012): attention switch indicated by P3a may, for

example, enable the evaluation of the relevance of rule violation with respect to learning. Also results suggesting that hippocampal lesions reduce P3a to novel sounds imply that P3a may be linked with learning (Polich, 2007; Friedman, Nessler, Kulik, & Hamberger, 2011). In the case of strongly established native-language rules, however, the updating of their memory representations on the basis of occasional illegal stimuli is likely rejected, and thus no long-term learning occurs.

In conclusion, phonologically illegal word forms elicited enhanced MMN and P3a responses. We suggest that these effects are induced by hierarchical predictive coding at two levels, which are based on predictions with respect to the sound sequence and phonological rules. Predictive coding of phonological rules could facilitate the online processing of speech by constraining the expected sound segments. Taken together, the present and previous findings suggest that the processing of continuous speech in natural communication is likely facilitated by hierarchically organised levels of predictive coding. These may include syntax (Pulvermüller et al., 2008), word recognition (Gagnepain et al. 2012; Ylinen et al., submitted), and phonological processing, as suggested by the present and previous studies (Hwang et al., 2010; Poeppel & Monahan, 2011; Scharinger et al., 2011, 2012; Weber, 2011). Investigating language learner's brain responses during the learning process could shed light on when and how these neural phonological predictions are formed.

Acknowledgements

This work has been funded by the Academy of Finland (project numbers 131963 and 274058) and Research Funds of the University of Helsinki (project number 76302111). The sponsors did not have any role in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

References

- Aaltonen, O., Hellström, Å., Peltola, M. S., Savela, J., Tamminen, H., & Lehtola, H. (2008). Brain responses reveal hardwired detection of native-language rule violations. *Neuroscience Letters*, *444*, 56-59.
- Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology*, *21*, 204.
- Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, *83*, 120-131.
- Bendixen, A., Scharinger, M., Strauss, A., & Obleser, J. (2014). Prediction in the service of comprehension: Modulated early brain responses to omitted speech segments. *Cortex*, *53*, 9-26.
- Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.45, retrieved 28 October 2010 from <http://www.praat.org/>.

- Bonte, M. L., Mitterer, H., Zelligui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, *116*, 2765-2774.
- Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibanez, A., Owen, A.M., & Bekinschtein, T.A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience*, *33*, 11194-11205.
- Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635-647.
- Escera, C., & Corral, M. (2007). Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *Journal of Psychophysiology*, *21*, 251.
- Friedman, D., Nessler, D., Kulik, J., & Hamberger, M. (2011). The brain's orienting response (novelty P3) in patients with unilateral temporal lobe resections. *Neuropsychologia*, *49*, 3474-3483.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815-836.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*, 293-301.
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, *22*, 615-621.

- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, *120*, 453-463.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*, 613-656.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*, 407-422.
- Hwang, S. K., Monahan, P. J., & Idsardi, W. J. (2010). Underspecification and asymmetries in voicing perception. *Phonology*, *27*, 205-224.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *23*, 9541-9546.
- Karlsson, F. (1983). *Suomen kielen äänne- ja muotorakenne*. Juva, Finland: WSOY.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*, 1-86.
- Mermelstein, P. & Fitch, H. (1976). Difference limens for formant frequencies of steady-state and consonant-bounded vowels. *The Journal of the Acoustical Society of America*, *60*, S119.
- Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, *65*, 956-969.

- Mitterer, H., Csepe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, 30, 451-479.
- Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, 109, 15953-15958.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1-21.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313-329.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544-2590.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357.
- Poeppel, D., & Monahan, P. J., (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, 26, 935-951.

- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128-2148.
- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J., & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, *14*, 607-616.
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*, 49-71.
- Pulvermüller, F., Shtyrov, Y., Hasting, A. S., & Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, *104*, 244-253.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79-87.
- Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., & Obleser, J. (2012). A sparse neural code for some speech sounds but not for others. *PloS One*, *7*, e40953.
- Scharinger, M., Idsardi, W. J., & Poe, S. (2011). Neuromagnetic reflections of harmony and constraint violations in Turkish. *Journal of Laboratory Phonology*, *2*, 99-124.
- Schröger, E., Marzecova, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*, 641-664.
- Steinberg, J., Jacobsen, T. K., & Jacobsen, T. (in press). Repair or violation detection? Pre-attentive processing strategies of phonotactic illegality demonstrated on the

constraint of g-deletion in German. *Journal of Speech, Language, and Hearing Research*.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2010a). Preattentive phonotactic processing as indexed by the mismatch negativity. *Journal of Cognitive Neuroscience*, 22, 2174-2185.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2010b). Activation and application of an obligatory phonotactic constraint in German during automatic speech processing is revealed by human event-related potentials. *International Journal of Psychophysiology*, 77, 13-20.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2011). Phonotactic constraint violations in German grammar are detected automatically in auditory speech processing: A human event-related potentials study. *Psychophysiology*, 48(9), 1208-1216.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2012). The role of stimulus cross-splicing in an event-related potentials study. Misleading formant transitions hinder automatic phonological processing. *Journal of the Acoustical Society of America*, 131, 3120-3140.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Eger, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*, 11, 1004-1006.

Sun, Y., Giavazzi, M., Adda-Decker, M., Barbosa, L. S., Kouider, S., Bachoud-Levi, A. C. et al. (2015). Complex linguistic rules modulate early auditory brain responses. *Brain and Language*, 149, 55-65.

- Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: A magnetoencephalography study. *PloS One*, *4*, e4452.
- Truckenbrodt, H., Steinberg, J., Jacobsen, T. K., & Jacobsen, T. (2014). Evidence for the role of german final devoicing in pre-attentive speech processing: A mismatch negativity study. *Frontiers in Psychology*, *5*, 1317.
- Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*, 3665-3678.
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*, 20754-20759.
- Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, *44*, 95-118.
- Wiik, K. (1965). *Finnish and English vowels*. (Annales Universitatis Turkuensis, Series B, Vol. 94). Turku, Finland: University of Turku.
- Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*, 147-163.
- Ylinen, S., Bosseler, A., Junttila, K., & Huotilainen, M. (submitted). Predictive coding mediates word recognition and learning in the early stages of language development.

- Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., Mäkelä, J.P., Service, E. (2015). Two distinct auditory-motor circuits for monitoring speech production as revealed by rehearsal-elicited suppression of auditory cortex. *Cerebral Cortex*, 25, 1576-1586.
- Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*, 1072, 175-185.

Figure legends

Figure 1. Stimulus waveforms. The cross-splicing of the stimuli was done during the silent occlusion phase of the stop consonant. The deviant stimulus violating the Finnish vowel harmony rule has been marked with an asterisk.

Figure 2. The grand-average ERPs for all auditory stimuli. The deviant stimulus violating the Finnish vowel harmony rule has been marked with an asterisk.

Figure 3. Deviant-minus-standard difference waveforms. The electrode is C3, where the largest MMN responses were observed. The deviant stimulus violating the vowel harmony rule of the Finnish language has been marked with an asterisk.





