Brain & Language 162 (2016) 72–80

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

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

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

ARTICLE INFO

Article history:
Received 26 November 2015
Revised 5 August 2016
Accepted 20 August 2016

Keywords:
Predictive coding
Prediction error
Phonological rules
Vowel harmony
Event-related potential (ERP)
Mismatch negativity (MMN)

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.

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, & Egner, 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, & 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 & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris & 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, Bendixen, Trujillo-Barreto, & Obleser, 2012; Scharinger, Idsardi, & Poe, 2011; Weber, 2001).

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 Bendixen et al., 2012;

* Corresponding author.
E-mail address: sari.ylinen@helsinki.fi (S. Ylinen).

http://dx.doi.org/10.1016/j.bandl.2016.08.007
0093-934X/

Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009; Wacongne, Changeux, & Dehaene, 2012; Wacongne et al., 2011). 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 (2016) 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, Truckenbrodt, & Jacobsen, 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., /tɑlo+ssɑ/ '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 [olympiɑ] 'Olympic'.

The Finnish vowel harmony has been previously studied with MMN by Aaltonen et al. (2008; see also Scharinger et al., 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 Finnic 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 /ɑ/ 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