Neural discrimination of speech sound changes in a variable context occurs irrespective of attention and explicit awareness

P. Virtala[a,b,*], E. Partanen[a,b,c], M. Tervaniemi[a,b,d], T. Kujala[a,b]

[a] Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland
[b] Cognitive Brain Research Unit, Institute for Behavioural Sciences, University of Helsinki, Finland
[c] Center of Functionally Integrative Neuroscience (CFIN), Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[d] Cicero Learning, University of Helsinki, Finland

ARTICLE INFO

ABSTRACT

To process complex stimuli like language, our auditory system must tolerate large acoustic variance, like speaker variability, and still be sensitive enough to discriminate between phonemes and to detect complex sound relationships in, e.g., prosodic cues. Our study determined discrimination of speech sounds in input mimicking natural speech variability, and detection of deviations in regular pitch relationships (rule violations) between speech sounds. We investigated the automaticity and the influence of attention and explicit awareness on these changes by recording the neurophysiological mismatch negativity (MMN) and P3a as well as task performance from 21 adults. The results showed neural discrimination of phonemes and rule violations as indicated by MMN and P3a, regardless of whether the sounds were attended or not, even when participants could not explicitly describe the rule. While small sample size precluded statistical analysis of some outcomes, we still found preliminary associations between the MMN amplitudes, task performance, and emerging explicit awareness of the rule. Our results highlight the automaticity of processing complex aspects of speech as a basis for the emerging conscious perception and explicit awareness of speech properties. While MMN operates at the implicit processing level, P3a appears to work at the borderline of implicit and explicit.

1. Introduction

Efficient implicit and explicit detection of rules and regularities in the auditory stream is a prerequisite for language encoding and learning. For example, in order to recognize a phoneme uttered by different speakers as the same phoneme, an individual has to categorize various acoustic signals with different physical properties as belonging to one phoneme class of a language. Therefore, automatic and efficient speech processing in the central nervous system requires rapid detection of relevant invariant speech features and tolerance for acoustic variance. This is evidenced by studies showing neural discrimination of native language speech sounds even when they vary widely acoustically due to different speakers, both in adults (Shestakova et al., 2002) and in newborn infants (Dehaene-Lambertz & Pena, 2001).

In studies of central auditory and speech processing, the mismatch negativity (MMN) and other change-related ERP components recorded with the electroencephalogram (EEG) and magnetoencephalogram (MEG) have been deemed informative research tools. The MMN is a fronto-central negativity elicited around 100–250 ms after deviance onset in a repetitive sound stream, and it has been used to study sensory memory functions, auditory change discrimination, and predictive coding in the auditory system from infancy to adulthood (Kujala, Tervaniemi, & Schröger, 2007; Näätänen, Kujala, & Winkler, 2011). According to recent models, the MMN is a component reflecting automatic change-discrimination of the human brain in conditions where the perceived sound differs from the auditory system's prediction of the expected auditory input (Kujala et al., 2007; Näätänen et al., 2011). The positive P3a component, which follows the MMN when the change in the sound stream is salient, is thought to reflect involuntary attention switching towards the stimulus and updating of working memory (e.g. Alho et al., 1998; Horváth, Winkler, & Bendixen, 2008).

Although the MMN is elicited independent of whether the listener attends to the sound stream or not (Näätänen et al., 2011; Sussman, 2007), it can reflect behavioral discrimination accuracy, i.e., MMN elicitation or increased MMN amplitude is associated with improved behavioral detection of changes in the sound stream (e.g. Amenedo & Escera, 2000; for review, see Kujala & Näätänen, 2010). In a study with simple duration changes varying in magnitude, MMN size was related to the magnitude of the deviance (the larger the deviance, the larger the MMN), and MMN size correlated positively with accuracy and

negatively with reaction time in a behavioral detection task (Amenedo & Escera, 2000; see also, Jaramillo, Paavilainen, & Näätänen, 2000; Novitski, Tervaniemi, Huotilainen, & Näätänen, 2004). Also P3a amplitude and latency can be sensitive to deviance magnitude and correlate with accuracy and reaction time of behavioral change detection (Novitski et al., 2004). This relationship with behavioral discrimination performance has made the MMN and P3a responses powerful tools to study auditory processing abilities and to compare groups with different motivational or attentional resources (for example clinical and healthy control groups) that would most likely contaminate any results obtained in behavioral discrimination tasks.

The relationship between the MMN and P3a and behavioral detection performance is, however, not always evident. MMN elicitation at the group level does not necessarily imply that individuals could behaviorally detect the deviance or verbally describe the change, meaning explicit awareness of its nature (Paavilainen, Arajärvi, & Takegata, 2007; van Zuijen, Simoens, Paavilainen, Näätänen, & Tervaniemi, 2006; for a recent review, see Paavilainen, 2013). For example, van Zuijen et al. (2006) recorded MMN and P3a responses to violations of a rule on frequency relationships (tone pairs descending in pitch, in the context of tone pairs ascending in pitch) in conditions where the sound stream was either ignored or attended. An MMN was elicited by the rule violations in a condition where the sound stream was to be ignored, even when the participants could not explicitly describe the violation after the condition. However, only participants who had explicit knowledge of the violation demonstrated a P3a response to the violations that they detected in a following attentive listening condition (van Zuijen et al., 2006).

There is evidence that the MMN can be elicited even when behavioral detection of the deviants remains at chance level, i.e., when not even implicit detection of deviants is evident (Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012; van Zuijen et al., 2006). Consequently, it has been suggested that appearance of the MMN in a learning task may even *predict* later increases in behavioral detection performance (Atienza, Cantero, & Dominguez-Marin, 2002; Tremblay, Kraus, & McGee, 1998). Thus, there seems to be a relationship between MMN elicitation and explicit awareness and behavioral detection of changes, but it is not straightforward (Paavilainen, 2013). However, direct comparisons of change-related ERPs with measures of behavioral detection performance and explicit awareness of changes are rare. Detailed understanding of the relationship of the MMN and P3a parameters (and elicitation), speed and accuracy of behavioral detection performance, as well as the ability to explicitly describe the heard changes are needed in order to understand the role of implicit and explicit levels of key perceptual processes, and efficiently utilize these ERPs in studies of auditory cognitive processing, e.g., in infants, children, and clinical populations.

MMN studies on the detection of complex rules and regularities in the auditory stream have so far used non-speech stimuli, despite the importance of these skills in speech processing. Originally, in a pioneering study, sound pairs with a falling pitch elicited MMNs when presented among sound pairs with a rising pitch in an ignore condition, even when sound pairs were presented at various frequency levels and no new frequencies appeared in the deviant sound pairs (Saarinen, Paavilainen, Schröger, Tervaniemi, & Näätänen, 1992). A further study demonstrated that these rule violations can be detected even when the interval width (pitch relationship) between sounds in a sound pair varies (Paavilainen, Jaramillo, Näätänen, & Winkler, 1999). This kind of rule extraction from auditory input is required in speech perception and language learning in order to, for example, master phonotactic rules of a language (Trask, 1996), or to detect phrase boundaries based on prosodic cues (e.g., Frazier, Carlson, & Clifton, 2006). Speech is a salient stimulus that attracts listeners' attention, while at the same time many aspects of speech are supposedly processed highly automatically: we speak, read and produce our native language intuitively, without explicit awareness of its rules and regularities. The automaticity, the

role of attention, and explicit awareness of the sound properties are thus intriguing phenomena in native language speech processing. However, this issue has not been systematically addressed in most previous studies on neural speech sound processing (e.g., Shestakova et al., 2002).

The present study investigated the automaticity of detecting phoneme changes and complex sound relationships in variable speech sound context, and the influence of attention on the detection process. To this end, pre-attentive, attentive, implicit, and explicit processing levels of native language phonemes with acoustic variation (F0-variation), mimicking variability in naturally occurring speech, were studied with ERP and behavioral experiments in healthy Finnish-speaking adults with no reading-, language-, or attention-related deficits. We determined how phoneme changes and violations of a rule on frequency relationships between sounds are neurally discriminated in unattended and attended sound streams. The tasks in attentive listening conditions included brief training of sound change detection, preceded and followed by queries about the participants' explicit awareness of the changes, and, after the queries, behavioral detection of sound changes.

Our specific research questions were as follows: (1) Do adults discriminate changes in native language phonemes and violations of abstract rules in a variable speech sound context at the pre-attentive and attentive neural levels and at the perceptual level? (2) How are neural and perceptual levels of change discrimination connected: Do neural response properties correlate with task performance, or can pre-attentive change discrimination be evident at the neural level even when change detection as measured by task performance is poor or non-existent? (3) What is the role of explicit awareness of deviance in pre-attentive neural auditory change discrimination: Are neural response properties associated with explicit awareness of deviances, and is pre-attentive change discrimination evident at the neural level even when explicit awareness of change types is poor or non-existent?

We hypothesized that, in line with previous studies, adults would differentiate between native language phonemes in a variable context effortlessly, and that this is reflected as MMN and P3a responses in pre-attentive (ignore) and attentive conditions as well as accurate performance in the behavioral detection task and good verbal descriptions of the deviancy. However, we expected that detecting violations of the abstract rule would be more challenging at all processing levels and gaining explicit awareness of it may require training. Particularly the P3a to the rule violation might be absent in the group data at least in the ignore condition. Furthermore, we hypothesized that the MMN and P3a properties are associated with behavioral detection performance as well as explicit awareness of the deviance: accurate and fast behavioral detection performance and explicit awareness of the deviance would result in larger and earlier responses, and the P3a would not be elicited without explicit awareness of deviance. Yet, we expected MMNs to be elicited irrespective of performance level in the behavioral tasks or explicit awareness of the deviance.

2. Methods

2.1. Participants

Altogether 21 participants, recruited via social media and web site of the current project, were included in the present experiment (9 male, mean age 26 years, range 20–36 years). Data of additional 2 individuals were excluded from analysis since one participant did not take part in all sessions of the study and the other one did not have a sufficient EEG data quality. All participants included were Finnish-speaking, right-handed and reported neither problems related to hearing or basic motor functions nor a history of language-, learning-, or attention-related difficulties or neurological disorders.

Of the participants, three reported that Italian or Swedish in addition to Finnish was spoken in their childhood home. All had attended a Finnish speaking school. In addition to their home languages, all

participants reported good or excellent skills in English, four reported good or excellent skills in another language (Swedish, French), and some reported limited skills in other languages (German, Russian, Spanish, Japanese, and Chinese). Ten of the 21 participants reported having some formal music education outside school (mean 5 years, range 1–10 years) and 20 reported listening to music on a regular basis, but no one reported having absolute pitch or being a musician. The participants had on average 16 years of education including elementary school (range 13–20 years). All participants gave a written informed consent to participate in the present study and received a compensation for their participation (vouchers for cultural or exercise activities) after completing the study. This study received ethical approval of the University of Helsinki Review Board in Humanities and Social and Behavioural Sciences.

2.2. Experimental stimuli and conditions

The auditory stimuli consisted of Finnish phonemes /i/ and /æ/ uttered by a female Finnish native speaker. Stimuli were edited using Praat 5.4.01 (Boersma & Weenink, 2013) and Adobe Audition CS6 5.0. Build 708 (Adobe Systems Inc., California, USA). These vowels were chosen for their large acoustic difference (Wiik, 1965) and because they sounded natural when their pitch was increased or decreased to create the stimuli for the experiment (see below). Sound intensity levels were RMS normalized between the stimuli, and the stimuli were cut from the end so that the duration of each phoneme was 230 ms. The natural attack of the stimuli was retained, while smooth fade out function in Adobe Audition was administered in the stimulus ending at 190–230 ms. The vowels had a natural F0-level of ~205 Hz, and they were transposed to seven additional frequency levels 1, 2, and 3 semitones lower and 1, 2, 3, and 4 semitones higher than the natural F0-level. As a result, both /i/ and /æ/ phonemes had eight different frequency levels in F0-range 172–258 Hz (174.3, 184.6, 195.4, 206.8, 217.8, 229.7, 242.5, and 256.2 Hz; typical range in female speech). The vowels from eight F0-levels were combined into vowel pairs /i/-/i/ or /i/-/æ/, so that all possible frequency combinations were generated, except that the pitch difference between the vowels within a pair was at least two semitones, and in /i/-/æ/ pairs, /æ/ always had a higher frequency than /i/ (thus, /æ/ phoneme was not presented at the two lowest frequency levels). As a result, 42 /i/-/i/ pairs (21 with a rising and 21 with a falling frequency) and 21 /i/-/æ/ pairs with different frequency combinations were generated. The duration of the vowel pairs was 530 ms, including a silent gap of 70 ms between the vowels.

The vowel pairs were presented in an oddball paradigm (Fig. 1), where /i/-/i/ pairs with a rising pitch (the first vowel with a lower frequency than the second vowel) occurred as the repeating standard stimulus with a probability of 80%. Vowel pairs /i/-/æ/ with a rising pitch, termed vowel deviants, and vowel pairs /i/-/i/ with a falling

pitch (the first vowel with a higher frequency than the second vowel), termed rule violations, acted as occasional deviant stimuli, each presented with a probability of 10%. Vowel pairs were presented in a pseudo-random order with the exception that at least one standard pair preceded every deviant pair. The time from the beginning of the vowel pair until the beginning of the next vowel pair was 1000 ms with a 25-ms-jitter in six 10-ms steps, resulting in 975, 985, 995, 1005, 1015, and 1025 ms between consecutive stimulus onsets to reduce phase-locked brain activity to regularly repeating stimuli.

Altogether four sequences with identical stimulus types and probabilities were constructed: a 21-min-long sequence (1) with 1260 stimuli (126 stimuli per deviant type) for the ignore/pre-training condition, a 90-s-long training sequence (2a) with 90 stimuli (9 stimuli per deviant type), and two 10-min-39-s-long sequences (2b and 2c) with 630 stimuli each (63 stimuli per deviant type) for the attentive/post-training condition.

2.3. Experimental procedure

The experiment was carried out in two sessions, separated by an optional break (see Table 1). During the experiment, the participant sat in a comfortable chair in a soundproof, electrically shielded chamber, while the stimulus sequences were presented via headphones (Sony Dynamic Stereo Headphones, MDR-7506) with an approximate sound level of 65 dB SPL(a). The stimuli were presented with Presentation software (NeuroBehavioral Systems Inc., California, USA), version 17.2. In the attentive conditions, behavioral responses were recorded with Cedrus RB844 response pad (Cedrus Corporation, California, USA).

The first session (ignore/pre-training) was an ignore condition, where EEG was recorded and sequence 1 was presented while the participant watched a self-chosen soundless subtitled DVD movie and was instructed to avoid moving or blinking and not to pay attention to the sounds. The participants were not informed that they would be queried about the stimulus sequence afterwards. Ignore condition was always presented first in order to avoid carry-over effects of attention and explicit awareness of the deviants. After this, EEG recording and DVD were paused and the participant was queried about the nature of the two deviants presented in the sound stream (based on what they remembered of the ignore condition).

The second session was a training phase (without EEG recording), where the participant was given a response pad and instructed to listen to the sound stream and press a button when he/she recognizes either of two changes in the auditory stream. Sequence 2a was presented only after ensuring that the participant understood the given task. After the training phase, the participant was again queried about the nature of the two presented deviants in the sound stream and the participant's verbal response was registered. It was then ensured that the participant understood the nature of the deviants, and it was rephrased and
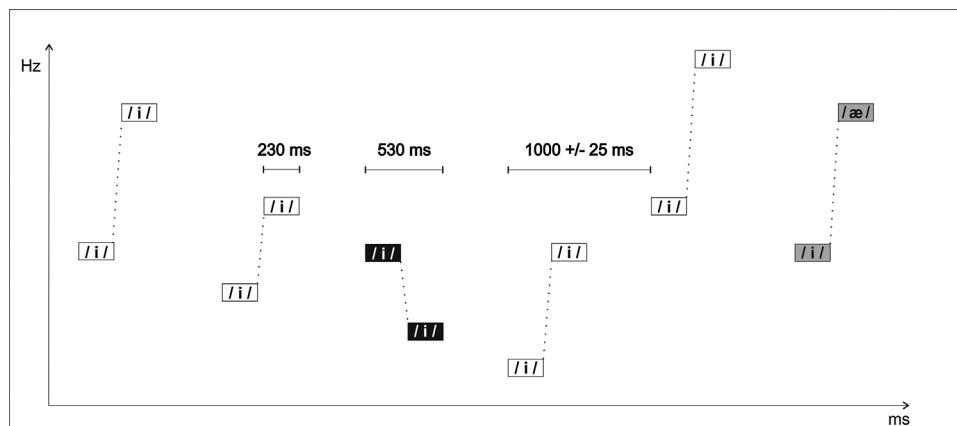

Fig. 1. The experimental condition, consisting of vowel pairs /i/-/i/ with a rising frequency (standards, probability 80%, white), vowel pairs /i/-/æ/ with a rising frequency (vowel deviants, 10%, grey), and vowel pairs /i/-/i/ with a falling frequency (rule violations, 10%, black).

Table 1
Experimental protocol. Instructions (I) were always given before the sequence, and queries (Q) and correct answers (A) after it. Sequences 2b and 2c were presented in a counterbalanced order between participants.

Sequence	EEG	Condition	Given instructions (I), correct answers (A), and queries (Q)
1	Yes	Ignore/ Pre-training	I: *"Your task is to focus on the silent movie and ignore the presented sound stream."* Q: *"The sound stream you heard varies constantly, but occasionally, two types of changes occur in it. Can you name these changes in the sound stream?"*
2a	No	Attentive/ Training	I: *"Next, focus on listening to the sound stream and think what changes occur in it. Push the response box button always immediately after hearing a change."* Q: *"Can you now tell, what kind of changes occur in the sound stream?"* A: *"In the stream of speech sounds, there are vowel changes (from/i/to/æ/) and rule violations as follows: when in most of the vowel pairs, pitch rises so that first sound is lower than the next sound, in some vowel pairs it falls, so that first sound is higher than the next sound."*
2b	Yes	Attentive/ Post-training	I: *"Again, focus on listening to the sound stream. Push the response box button always immediately after hearing the vowel change. Ignore the rule violation."*
2c	Yes	Attentive/ Post-training	I: *"Again, focus on listening to the sound stream. Push the response box button always immediately after hearing the rule violation. Ignore the vowel change."*

repeated until the participant indicated having understood it. In the third session (attentive/post-training), in a counterbalanced order between participants, the participant was given the instruction to detect first one of the deviant types (vowel/rule) and then the other one in two consecutive sequences 2b and 2c while his/her EEG was recorded.

2.4. EEG recording and analysis

The EEG was recorded continuously with 64 electrodes (headcap and amplifier: Biosemi ActiveTwo mk2, BioSemi B. V., Amsterdam, The Netherlands) placed according to the international 10–20-system, with additional 5 external active Ag/AgCl electrodes (right and left mastoid behind the ears, vertical and horizontal electro-oculogram below and next to participant's left eye, tip of the nose) with an online sampling rate of 512 Hz. EEG was imported to the BESA analysis program (v 6.0, BESA GmbH, Gräfelfing, Germany), filtered at 1–30 Hz (slope 12 dB/oct., zero phase) and re-referenced to the mean of the mastoid electrodes. Automatic eye artifact correction was conducted (detection thresholds 150 μV for horizontal and 250 μV for vertical eye movements, v 6.0, BESA, Berg & Scherg, 1994). Channels in the periphery (close to the edges of the EEG cap) where the EEG signal contained a lot of noise by visual inspection were excluded from the data. When bad channels were not located in the periphery, they were interpolated (max 3 channels interpolated per participant per condition). The data were divided to epochs individually for each vowel pair (-100 to 975 ms from the vowel pair onset), and averaged separately for each participant, stimulus type (standard for vowel deviant, vowel deviant, standard for rule deviant, rule deviant), and electrode in each condition of the experiment. All epochs with voltage changes exceeding $\pm$ 120 μV were omitted from further analysis. A baseline correction for -100 to 0 ms was applied to all epochs prior to statistical testing.

In order to study the responses related to deviance processing, subtraction curves were calculated individually for each participant, stimulus type, electrode, and condition, so that the ERP waveform in response to the standard stimulus was subtracted from the ERP waveform in response to the deviant stimulus separately for each deviant in conditions 1, 2b, and 2c. In the attentive conditions, ERPs in response to attended but not target deviants (e.g., vowel deviants in 2c) were calculated, as targets were expected to contaminate the EEG signal due to motor activity related to button presses.

As rule deviants had a falling pitch, the first tone in the rule deviants could not be from the two lowest F0 levels (as the second tone should be at least two F0 levels lower than the first). Similarly, as standards had a rising pitch, the first tone in the standards could not be from the two highest F0-levels. Thus, when the first tone in the tone pair had one of the mentioned F0 levels, the auditory system might implicitly notice that the tone pair was a standard or a rule deviant on the basis of the first tone only. In order to avoid change-related ERPs to these stimuli, only the standards and rule deviants starting from the four middle F0-

levels were included when constructing the average ERPs. However, when constructing the average ERPs and subtraction curves for the vowel deviants, all standard and vowel deviant tone pairs were included to the averages (as the first tone did not reveal anything about the following tone). As a result, different standard responses were used to study rule and vowel detection. In the final dataset, all accepted participants had at least 80% of accepted trials per stimulus type in each condition. On average, one participant had 124 (range 114–126) accepted vowel deviant trials and 59 (51–60) accepted rule deviant trials in ignore condition and 62 (57–63) vowel deviant trials and 29 (24–30) rule deviant trials in each attentive condition. Data from one condition (2c) was missing from one participant due to a technical problem in data acquisition.

For the statistical analysis, peak latencies of the MMN and P3a responses were searched from time windows defined according to prior studies and visual inspection of the waveforms (100–300 ms for the MMN, 150–450 ms for the P3a,[1] starting from the onset of the 2nd vowel of the vowel pair) at the frontal midline electrodes Fz, FCz, and Cz, where the responses usually demonstrate largest amplitudes (Kujala et al., 2007), individually for each participant, deviant, and condition. Differences in peak latencies between ignore and attentive conditions on the three midline electrodes Fz, FCz, and Cz were analyzed with repeated measures analyses of variance (ANOVA-Rs with within-subject's factors "electrode", 3 levels, and "attention", 2 levels, conducted separately for MMN and P3a peak latencies). MMN and P3a mean amplitudes were then calculated from 50-ms time windows centered at the group-mean peak latencies of each deviant and condition.

Statistical significance of the MMN and P3a mean amplitudes in response to vowel deviants and rule violations in the ignore and attentive conditions was analyzed at Fz (one-sample two-tailed *t*-tests of the mean amplitudes against 0). A region-of-interest of 9 electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) was additionally used to study the effect of attention on mean amplitudes of the responses by comparing the ignore and attentive conditions (ANOVA-R with within-subject's factors "electrode", 9 levels, and "attention", 2 levels, conducted separately for MMN and P3a mean amplitudes). In the ANOVAs, differences between electrodes or interactions between electrodes and attention were not investigated. Greenhouse-Geisser correction was applied when the assumption of sphericity was violated. When a statistically significant factor had more than two levels, post-hoc comparisons were conducted, and a Bonferroni correction for multiple testing was applied. None of the electrodes used in the analyses were excluded or interpolated in any participant.

[1] For P3a to vowel deviant in the attentive condition, peak latency was searched from a narrower time window of 150–350 ms, as a second positive peak appeared in the grand average waveform around 400 ms post-deviance (see Fig. 2).

Table 2

The MMN and P3a peak latencies and mean amplitudes with standard deviations (sd) in ignore and attentive conditions for the two deviant types on Fz electrode. Mean amplitudes differing statistically significantly from 0 are marked with asterisks.

Vowel	Condition	Time window (ms)[a]	Peak latency (ms)		Mean amplitude (μV)	
			mean	sd	mean	sd
MMN						
	Ignore	115–165	142.3	37.1	−2.43***	1.18
	Attentive	135–185	161.3	38.6	−3.46***	1.99
P3a						
	Ignore	200–250	226.0	12.4	3.76***	1.39
	Attentive	225–275	252.7	32.0	3.05***	2.38

Rule	Condition	Time window (ms)	Peak latency (ms)		Mean amplitude (μV)	
			mean	sd	mean	sd
MMN						
	Ignore	180–230	207.3	36.8	−0.80***	0.88
	Attentive	190–240	217.9	43.5	−1.08**	1.21
P3a						
	Ignore	300–350	332.8	56.0	0.58**	0.72
	Attentive	350–400	375.6	57.0	1.15**	1.31

in one-sample *t*-tests against 0. *** p < 0.001. ** p < 0.01. * p < 0.05.

[a] all latencies are reported from the beginning of the second vowel of the vowel pair (300 ms from vowel pair onset).

2.5. Analysis of behavioral data

Verbal responses to the queries after the ignore condition and after training were scored from 0 to 2 for each deviant, 0 indicating no explicit awareness of the nature of the deviant, 1 indicating some explicit awareness (e.g., that phonemes/vowels changed in the vowel deviant or that the sound order or relationships or "melody" changed in the rule violation), and 2 indicating complete explicit awareness (e.g., that vowel changed to/æ/in the vowel deviant or sound pairs/groups had a falling instead of a rising pitch in the rule violation).[2] Response scores after ignore condition and after training were compared with non-parametric related-samples Wilcoxon signed ranked tests, as the distributions of the response scores were highly skewed (see Table 3).

For the behavioral results in the attentive conditions, the percentage of hits per button presses (hit-ratio) as well as reaction times were calculated for each individual separately for the training sequence (2a) and for the vowel (2b) and rule detection (2c) post-training conditions. As good task performance should be both accurate and fast, we expected good performance to be reflected as high percentage of hits per button presses (accuracy) and short reaction times (speed). For the training phase, the hit-ratio could not be calculated in the same way as in the other conditions, as the participants reacted to both deviant types at the same time. In order to still estimate the detection performance per deviant type during training, a hit-ratio was calculated for the amount of hits per deviant in relation to the amount of button presses that were not hits to the other deviant:

$$\text{ratio vowel} = \frac{\text{hits vowel}}{\text{button presses} - \text{hits rule}}$$

$$\text{ratio rule} = \frac{\text{hits rule}}{\text{button presses} - \text{hits vowel}}$$

Hit-ratios for each deviant type in the training and post-training conditions were compared against chance-level with one-sample *t*-tests. As the probability of a deviant was 10% in sequences 2b and 2c, the likelihood of hitting a deviant by chance on each button press was 10%.

The relationship between explicit awareness of deviants and behavioral detection performance was studied by grouping the participants based on their explicit awareness of the deviants in different phases of the experiment (after ignore condition, after training). One-sample *t*-tests were conducted to analyze whether the hit-ratios were above chance-level in each group. Hit-ratios and reaction times between the groups were compared with non-parametric independent samples Mann-Whitney *U* tests, as many of the hit-ratios and reaction times were not normally distributed [hit-ratio for vowels in post-training condition, Shapiro-Wilk(18) = 0.58, p < 0.001; reaction time for rule detection during training, Shapiro-Wilk(18) = 0.79, p < 0.01; hit-ratio for rule detection during training, Shapiro-Wilk(18) = 0.88, p < 0.05)].
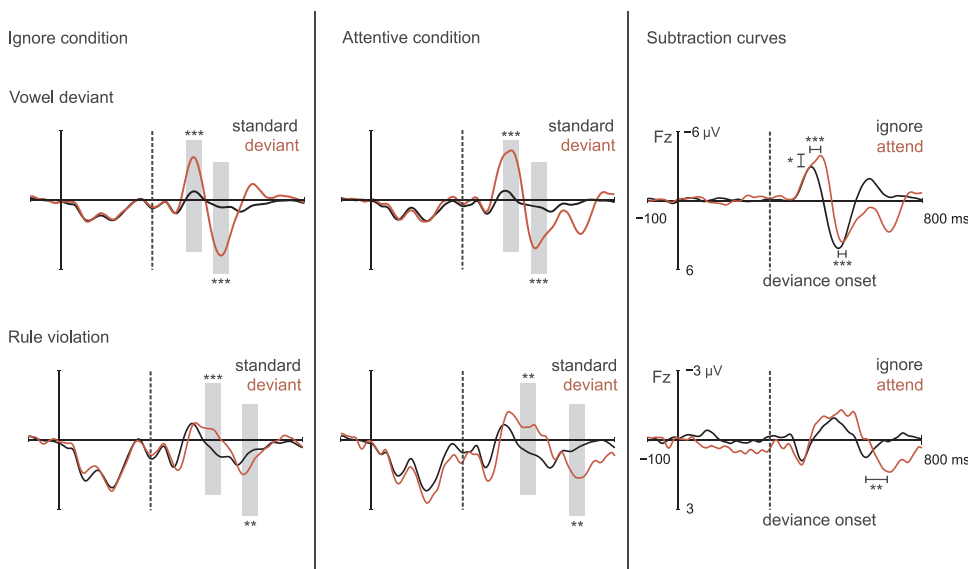
2.6. Analysis of the relationships of EEG and behavioral data

The relationship between explicit awareness of deviants and neural change detection was studied by grouping the participants based on their explicit awareness of the deviants in different phases of the experiment (after ignore condition, after training) and comparing the ERP properties in the ignore condition between the groups. All participants were informed about the nature of the deviants before the attentive EEG condition, and thus the participants did not differ in terms of their explicit awareness of the deviants in that phase of the experiment. Therefore the effect of explicit awareness on the ERPs in the attentive condition was not analyzed.

The analyses were conducted as follows: First, MMN and P3a elicitation in the groups in ignore condition was analyzed with one-sample *t*-tests. Then, group differences in the MMN and P3a amplitudes and peak latencies in ignore condition were analyzed with one-way ANOVAs. The relationships between task performance and ERPs in both ignore and attentive conditions were analyzed with non-parametrical Spearman correlation analyses, as many of the hit-ratios and reaction times were not normally distributed (see above). Associations between the MMN and P3a amplitudes and latencies *vs.* hit-ratios and reaction times in the different conditions were studied. In the Spearman analyses, the false discovery rate (FDR) correction was used to correct for multiple comparisons.

---

[2] Some participants reported hearing the /i/-/i/ vowel pairs as two different vowels, e.g., /y/-/i/, and thus they described the rule as an order change of the vowels, e.g., to /i/-/y/. As this resulted in accurate identification of the rule violations, this description was scored as a correct response.

Table 3

Verbal response scores. Amounts of participants who demonstrated no awareness (0 points), some awareness (1 point), and complete awareness (2 points) of the deviant types after the ignore condition and after training.

	0 points	1 point	2 points
vowel after ignore	5	4	12
vowel after training	2	6	13
rule after ignore	17	1	3
rule after training	10	2	9

3.2. Explicit awareness of the deviants and their behavioral detection

Most participants could verbally describe the vowel deviant already after the ignore condition and at least after training, indicating explicit awareness of the vowel deviant (Table 3). Vowel deviants were detected above chance both during training and in the post-training condition (Table 4). During training, one participant performed below chance-level in vowel detection, while all other participants and all participants in the post-training condition performed above chance.
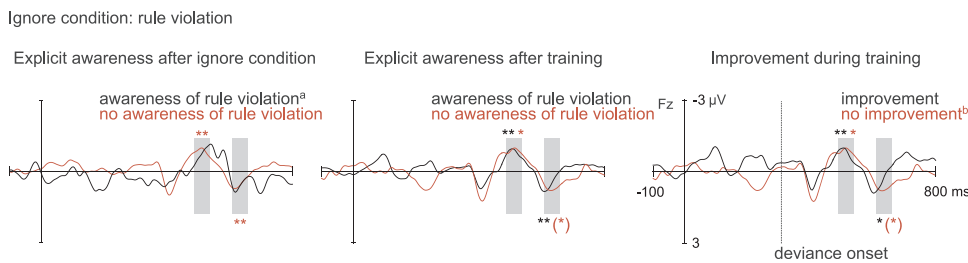

Fig. 2. ERPs to standard and deviant stimuli in ignore (left) and attentive condition (middle), and subtraction curves in ignore and attentive conditions (right, standard subtracted from deviant) on Fz. ERPs to vowel deviants are depicted on the top and those to rule violations on the bottom. Grey bars mark the time windows for MMN (earlier bar) and P3a (later bar) mean amplitudes, with statistical significances between the responses to the standard and the deviant (two leftmost panels) or between amplitudes (vertical bars, rightmost panels) or latencies (horizontal bars, rightmost panels) in ignore and attentive conditions marked with asterisks: *** p < 0.001, ** p < 0.01, * p < 0.05.

3. Results

3.1. ERPs in ignore and attentive conditions

The vowel deviant elicited statistically significant MMN and P3a responses in both ignore/pre-training and attentive/post-training conditions (Table 2, Figs. 2 and 4). Both the MMN and the P3a peaked earlier when the sounds were ignored than when they were attended: for the MMN, $F(1, 19) = 19.77$. $p < 0.001$, $\eta_p^2 = 0.51$, $\eta^2 = 0.24$, for the P3a, $F(1.19) = 13.30$, $p < 0.01$, $\eta_p^2 = 0.41$, $\eta^2 = 0.29$ (Table 2). The MMN to the vowel deviant was smaller when the sounds were ignored than when they were attended, $F(1, 19) = 6.23$, $p < 0.05$, $\eta_p^2 = 0.25$, $\eta^2 = 0.22$ whereas the P3a amplitude was not influenced by attention ($p > 0.10$).

Rule violations elicited statistically significant MMN and P3a responses in both ignore/pre-training and attentive/post-training conditions (Table 2, Figs. 2 and 4). The P3a peaked earlier when the sounds were ignored than when they were attended, $F(1, 20) = 9.97$, $p < 0.01$, $\eta_p^2 = 0.33$, $\eta^2 = 0.21$, whereas the MMN peak latencies and the MMN and P3a amplitudes were not influenced by attention ($p > 0.10$; for P3a amplitude $p = .10$).

Most participants were not able to verbally describe the rule violation after the ignore condition. However, explicit awareness of the rule violation increased as a result of training (related-samples Wilcoxon signed ranked test for verbal response scores after ignore condition *vs.* after training, $p < 0.05$). Rule violations were detected above chance both during training and in the post-training condition (Table 4). During training, three participants performed below chance-level in rule violation detection, while all participants in the post-training condition performed above chance.

Hit-ratios seemed to be higher and reaction times shorter in the post-training condition than during training for both the vowel deviants and the rule violations, but they were not compared statistically due to their different duration and type of task.

The task performance during training was not associated with explicit awareness of the vowel deviant either after ignore condition or after training (hit-ratio and reaction time, in all $p > 0.10$, see Table 5). In the training sequence, the hit-ratio was above chance-level in the group with some/good explicit awareness of the vowel deviant after the ignore condition and after the training sequence. All participants with no explicit awareness of the vowel deviant after the ignore condition or after training had hit-ratios above chance-level in the training

Fig. 3. ERP subtraction curves to rule violations in ignore condition in participants with some/good or no explicit awareness of the rule after ignore condition (left) or after training (middle), and in participants with improved or not improved explicit awareness of the rule as a result of training (right). Grey bars mark the time windows for MMN (earlier bar) and P3a (later bar) mean amplitudes, with their statistical significances marked with asterisks: ** p < 0.01, * p < 0.05, (*) 0.05 < p < 0.10. [a] No statistical testing as the group only had four

participants [b] The group is identical to the middle column "no explicit awareness" group.
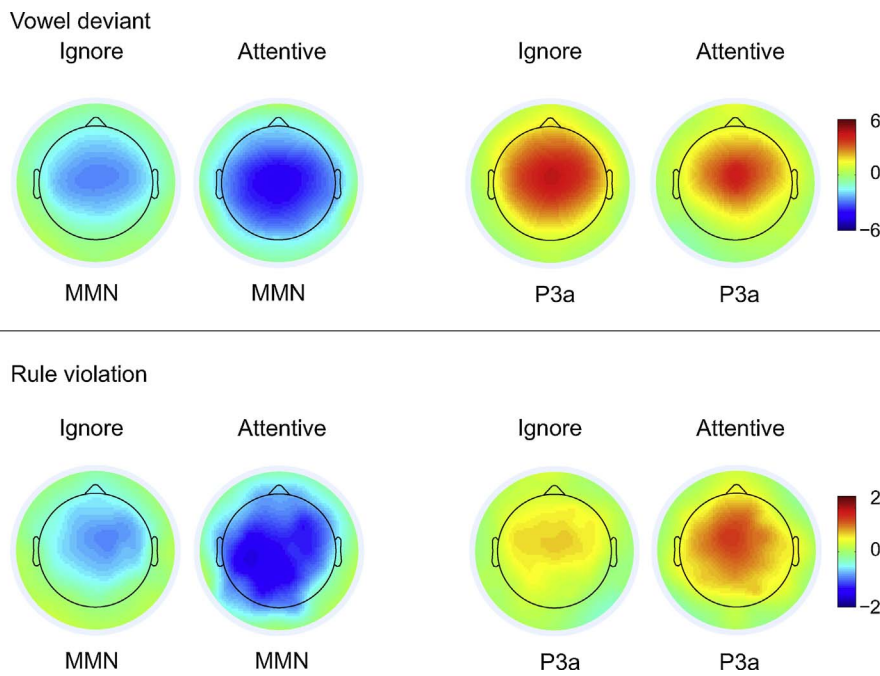
Vowel deviant

Fig. 4. Voltage maps of MMN and P3a mean amplitudes on the scalp in response to the vowel deviant (top) and the rule violation (bottom).

Table 4
Mean hit-ratios and reaction times during training and post-training (standard deviation in parentheses). Hit-ratios differing statistically significantly from chance-level (10%) are marked with asterisks.

	hit-ratio, %	range hit-ratio, %	reaction time, ms	range reaction time, ms
training: vowel detection	62.9 (27.9)***,a	6.0–100.0	532.8 (156.3)	361.0–959.6
post-training: vowel detection	93.2 (14.3)***	46.2–100.0	459.6 (85.5)	325.5–664.7
training: rule detection	43.1 (33.2)***,a	0.0–100.0	630.5 (268.3)	391.8–1312.9
post-training: rule detection	63.4 (25.4)***	23.5–100.0	549.0 (101.3)	342.0–741.0

[a] Calculated differently than other hit-ratios, see Methods.

*** $p < 0.001$ in one-sample t-tests against chance-level (10.0).

sequence.

The task performance in the training sequence was not associated with explicit awareness of the rule violation after the ignore condition (hit-ratio and reaction time, in both $p > 0.10$). All participants with some/good explicit awareness of the rule after ignore condition had hit-ratios above chance-level, and so had the group with no explicit awareness of the rule after the ignore condition. However, explicit awareness of the rule violation after training was associated with task performance in the training sequence (hit-rate, $p < 0.001$; reaction time, $p < 0.05$): hit-rate was higher and reaction time shorter in the training sequence in participants who had some/good explicit awareness compared to those who demonstrated no explicit awareness of the rule violation afterwards. Hit-ratio was above chance-level in the training sequence in participants who had some/good explicit awareness of the rule violation, but not in the group with no explicit awareness of the rule violation after training. All the three participants with below-chance hit-ratios for rule violations in the training sequence were in this group. Participants who demonstrated no awareness of the rule violation after the ignore condition, but whose explicit awareness improved as a result of training, had an above-chance rule violation hit-ratio during training.

Table 5
Amounts of participants (total n = 21) demonstrating no (verbal response score 0) or some/good explicit awareness (verbal response score 1 or 2) of the vowel change and rule violation after the ignore condition and after training, cross-tabulated with their behavioral performance in the following training sequence.

vowel deviant	mean hit-rate, %	mean reaction time, ms	total
after ignore/pre-training			
no explicit awareness	70.8	567.1	5
some/good explicit awareness	60.4***	522.1	16
after training			
no explicit awareness	62.5	544.1	2
some/good explicit awareness	63.0***	531.6	19

rule violation	mean hit-rate, %	mean reaction time, ms	total
after ignore/pre-training			
no explicit awareness	40.7**	659.8	17
some/good explicit awareness	53.3	527.9	4
after training			
no explicit awareness	16.3ns	786.4	9a
some/good explicit awareness	62.4**	545.6	11
improversb	68.0**	564.8	8

ns $p > 0.10$, ** $p < 0.01$, *** $p < 0.001$ in one-sample t-tests against chance-level.

[a] One participant demonstrated some awareness of the rule violation after ignore condition but no longer after training – he/she was not included in either group.

[b] Participants who had no explicit awareness of the rule after ignore condition but some/good explicit awareness of the rule after training.

3.3. Relationship between ERPs and explicit awareness of deviants

The participants who demonstrated no explicit awareness of the rule violation after the ignore condition (Table 5, Fig. 3), still showed statistically significant MMN and P3a responses to the rule violations in the ignore condition [for the MMN: $t(16) = -4.16$, $p < 0.01$; for the P3a: $t(16) = 3.18$, $p < 0.01$]. The participants who continued to demonstrate no explicit awareness of the rule violation even after training, still showed statistically significant MMN and nearly

significant P3a responses to the rule violations in the ignore condition [for the MMN: t(8) = −2.54, p < 0.05; for the P3a: t(8) = 2.17, p = .06]. The participants who demonstrated some/good explicit awareness of the rule violation after training also showed statistically significant MMN and P3a responses to the rule violations in the ignore condition [for the MMN: t(10) = −3.21, p < 0.01; for the P3a: t (10) = 3.44, p < 0.01]. These two groups did not differ in mean amplitudes or peak latencies of the MMN and P3a responses to the rule violations in the ignore condition (in all p > 0.10).

Furthermore, the participants who had no explicit awareness of the rule violation after ignore condition but whose awareness improved as a result of training (improvers, Fig. 3, right), also demonstrated statistically significant MMN, t(7) = −3.57, p < 0.01, and P3a responses, t (7) = 2.40, p < 0.05 in the ignore condition, with no statistically significant differences in the MMN or P3a amplitude or latency between the groups which improved vs. did not improve (in all p > 0.10). As only some individual participants demonstrated no explicit awareness of the vowel deviant (Table 5) and the vowel deviant was expected to be highly salient for all the participants, the effect of explicit awareness on the MMN and P3a responses to vowel deviant was not analyzed further.

3.4. The relationship between neural and behavioral change discrimination

In order to study the relationships between neural change-discrimination and behavioral detection performance, subtraction curves at Fz, FCz, and Cz electrodes were averaged together in order to improve the signal-to-noise ratio and the MMN and P3a mean amplitudes and peak latencies were re-calculated from the average curves. The hit-ratio for the vowel in the post-training condition statistically significantly correlated with the amplitude of the MMN to the vowel deviant in the attentive condition: the higher the hit-ratio, the larger the MMN (Table 6, Supplementary Fig. S1 in the online version at DOI: 10. 1016/j.biopsycho.2018.01.002). The MMN latency for the vowel deviant also had significant or nearly significant correlations with vowel detection reaction time both during training (MMN in ignore condition, p = 0.09) and in the post-training condition (MMN in ignore condition, p < 0.05), so that later latencies predicted shorter reaction times.

The hit-ratio for rule violations during training statistically significantly correlated with the MMN amplitude to rule violations in the attentive condition: the higher the hit-ratio, the larger the MMN. The reaction time for rule detection in the post-training condition correlated with the P3a amplitude to rule violations in the attentive condition: the longer the reaction time, the larger the P3a. However, when correcting for multiple comparisons, none of the found correlations remained statistically significant.

4. Discussion

The aim of the present study was to determine the automaticity and influence of attention on detecting phoneme changes and complex relationships between speech sounds in a variable context, mimicking speaker and prosodic variability in naturally occurring speech. We investigated phoneme discrimination and detection of abstract rule violations in speech sounds with acoustic variation to assess how neural processes are associated with task performance, and to study how the direction of attention as well as explicit awareness of the phoneme changes and rule violations influence these phenomena.

Phoneme changes and rule violations were neurally and behaviorally discriminated as indicated by MMNs both to vowel deviants and rule violations in ignore and attentive listening conditions, as well as their above-chance behavioral detection performance. Both deviants even elicited P3a responses in ignore and attentive listening conditions, suggesting that the changes were salient enough to involuntarily catch the participants' attention. These results support the notion of high automaticity in the discrimination of speech sound changes in a

variable context (Dehaene-Lambertz & Pena, 2001; Shestakova et al., 2002). Importantly, unlike previous studies, the present study systematically compared pre-attentive, attentive, implicit, and explicit, neural and behavioral processing levels of acoustically variable speech sounds. The results obtained suggest that also complex sound relationships are detected even without attentional effort in a speech sound context, thus extending previous work conducted with simpler stimuli (e.g., Paavilainen et al., 1999; van Zuijen et al., 2006). Furthermore, our results illuminate the long-debated automaticity of the MMN elicitation with respect to attention (Sussman, 2007) and the role of P3a in the chain of events that results in conscious perception of a change (Horváth et al., 2008).

In line with our hypotheses and previous work (e.g., van Zuijen et al., 2006), the MMN was elicited in the ignore listening condition irrespective of explicit awareness of the change properties. Crucially, even in participants who did not succeed in gaining explicit awareness of the rule violation during a short period of attentive listening, an MMN was elicited. Also P3a was elicited even in the absence of explicit awareness of the rule violation, but in participants who did not improve in their explicit awareness of the change, it was only nearly statistically significant. As expected, gaining explicit awareness of the rule violation mostly required attentive listening whereas participants effortlessly became aware of the vowel deviant. Even though the participants gave generally poor or inadequate descriptions of the rule violations prior to training, rule violations were still detected above chance by the majority of the participants in the training sequence and in the post-training conditions following it. As hypothesized, neural and behavioral change discrimination were related, as MMN amplitudes correlated with behavioral detection accuracy (in line with previous work, e.g., Amenedo & Escera, 2000; Novitski et al., 2004). However, this was only true for MMN responses in the attentive listening condition, and relationships of MMN peak latency and P3a amplitude and latency with task performance were less consistent. Furthermore, none of the correlations between ERP properties and task performance survived correction for multiple comparisons, possibly due to the small number of participants and large amount of tests conducted. Behavioral change detection was further associated with explicit awareness of the changes, as accurate and fast behavioral detection of rule violations during training predicted explicit awareness of the rule violation afterwards (in line with van Zuijen et al., 2006).

Taken together, the present results offer strong evidence for the automaticity of the neural discrimination process reflected by the MMN, while also demonstrating its link to the perceptual level of behavioral change detection. They illuminate the chain of neural events ranging from automatic sound discrimination to the conscious level of sound perception. In future studies, a larger sample size should be used to gain more statistical power and to have the possibility to form subgroups of participants to investigate the relationship between ERPs and perception.

4.1. Automaticity of rule violation discrimination

In the present study, explicit awareness of the rule violation was not a prerequisite for the MMN elicitation in the ignore condition. Participants who were not able to verbally describe the sound changes and thus demonstrated no explicit awareness of the rule violation after the ignore condition nevertheless had a statistically significant MMN response to the rule violations in the ignore condition. Even those participants who did not improve in their explicit awareness of the rule, i.e., who demonstrated no explicit awareness of the rule even after training, had a statistically significant MMN in the ignore condition at the group level. This was evident even though the sample sizes were small in these analyses, indicating that MMNs were robustly elicited in the participants.

Furthermore, the MMN responses were of similar magnitude and latency in participants who demonstrated no *vs.* some/good explicit

awareness of the rule violation. It is possible that the groups did not differ statistically because of a lack of power in the analysis − however, the result suggests that there were no large differences between the groups, supported by visual inspection showing that the MMNs also looked very similar in the two groups (Fig. 3). In the future, larger sample sizes should be included in order to make a more comprehensive analysis of subgroups: those who have explicit awareness of the deviants, those who do not, and those who succeed or fail in gaining explicit awareness of the deviants during the experiment.

The present MMN results are in line with previous findings with simpler, non-speech stimuli, showing pre-attentive MMN elicitation irrespective of whether the participants gained explicit awareness of the nature of the deviance during the experiment (Paavilainen et al., 2007; van Zuijen et al., 2006). A more recent study showed that participants had MMN responses to changes in tone sequences in an ignore condition, in spite of having chance-level performance in a following behavioral detection task of the same changes, which suggests no explicit awareness of the nature of the changes (Paraskevopoulos et al., 2012; see also Bendixen, Roeber, & Schröger, 2007; Bendixen & Schröger, 2008; Lang & Kotchoubey, 2000; Paavilainen, Simola, Jaramillo, Näätänen, & Winkler, 2001; Schröger et al., 2007). These converging results highlight the automaticity of auditory rule analysis, which might be a prerequisite for rapid and efficient perceptual processes.

A previous learning experiment with synthesized speech sound contrasts showed MMN changes in some participants before they mastered the speech sound contrast behaviorally (Tremblay et al., 1998). The authors suggested that MMN elicitation may precede, or coincide with, but not follow the emerging behavioral mastering of the change. In the context of the present study, this would suggest that even the participants who demonstrated no explicit awareness of the rule after training and had chance-level task performance during training, still had the potential to learn to master the rule, as indicated by MMN elicitation in the ignore condition. In fact, in line with this hypothesis, when the rule was explained to them, all participants did demonstrate above-chance behavioral detection. A more careful investigation of MMN elicitation *vs.* explicit awareness of the changes and behavioral change detection would require the analysis of MMN elicitation at the individual participant level. To this end, a separate study with a larger amount of stimuli, yielding a sufficiently high signal-to-noise ratio, should be carried out in the future.

In the present study, the P3a was also, although somewhat less consistently than the MMN, elicited in the absence of explicit awareness of the rule violation. Participants with no explicit awareness of the rule violation after the ignore condition had statistically significant P3a's to rule violations in the ignore condition. However, in the non-improving participants, the P3a was only nearly statistically significant and seemed smaller than in the improving group, although lack of significance could be attributable to the small sample size as the group difference in the P3a latency and amplitude did not reach significance. In general, this finding gives support to the three-stage model of auditory cognition, based on original findings in which P3a could be elicited without concurrent N1-increase or MMN elicitation and also without consequent elicitation of the re-orienting negativity RON (Horváth et al., 2008). In other words, these ERP indices of auditory cognition and distraction do not form a unified sequence but, instead, can appear irrespectively of each other.

In a similar vein, in a previous study, the P3a was elicited only when explicit awareness of the deviants was gained (van Zuijen et al., 2006). In contrast, in another study investigating the detection of violations in abstract rules, the P3a responses were elicited by the rule violations in a condition where the sound stream was to be ignored, and behavioral detection of the rule violations was somewhat above chance-level in a condition where the sounds were attended, even though the participants could not verbally describe the rules (Paavilainen et al., 2007). Together with these previous findings, the present results suggest that while the MMN is clearly operational at the implicit processing level of

sound change detection, the P3a seems to work at the borderline of the implicit and the explicit, being elicited not completely irrespective of explicit awareness of the changes but also not always requiring it. Thus, the MMN and the P3a seem to be essential but not identical steps in the chain of events resulting in a conscious experience of our surroundings.

4.2. Attention effects on neural speech sound processing

Whereas change-related ERPs to the vowel deviants and rule violations were elicited irrespective of attention, attention did have an influence on the amplitudes and latencies of these responses. The MMN to the vowel deviant peaked later and its amplitude was larger when the sounds were attended than when they were ignored. This can be an attention effect on the MMN, but it may also reflect the emergence of an overlapping attention-related N2b component when the sounds are attended, resulting in a later response with a larger amplitude than the "purer" MMN to unattended stimuli (Potts, Dien, Hartry-Speiser, McDougal, & Tucker, 1998; Ritter et al., 1992). As the vowel deviants were highly salient and explicitly detected even in the ignore condition, N2b contribution in response to the vowel deviant is likely in the attentive condition even though the task was to detect the rule violations. In contrast, as the rule violations were less salient and poorly detected by the participants, N2b contribution is less likely in response to them. Accordingly, the MMN to the rule violations was not influenced by attention in the present study. The results may thus be in concert with the suggestion by Sussman (2007): the MMN per se would not be affected by attention, but it can be overlapped by the N2b when the deviant is detected by the listener.

A small, but statistically significant P3a was elicited by the rule violations not only in the attentive condition but also in the ignore condition, in contrast with our hypothesis above that the rule violation would not attract attention in the ignore condition. It is notable that even though the rule violation was quite subtle, most of the participants could still behaviorally detect the rule violations in the training sequence following the ignore condition, and some individuals could even verbally describe the change immediately after the ignore condition. Occasional attention switches towards the rule violations may have happened at least in some participants during the ignore condition. In the present study, the P3a to both the vowel deviant and the rule violation peaked later when the sounds were attended than when they were ignored, while the P3a amplitude was not influenced by attention. This delay may have been caused mainly by intervening attention-related components (like the N2b) following the MMN. The interpretation of the P3a responses would benefit from, e.g., source modeling in future studies.

4.3. Association of ERPs and perceptual sound processing

In the present study, correlations were seen between ERP properties and task performance, but none of the found correlations survived statistical correction for multiple comparisons and should therefore be treated with strong caution. It is notable that also prior studies in the field have typically not corrected for multiple comparisons even when a large amount of correlation analyses were conducted (e.g., Novitski et al., 2004), which is a challenge in terms of statistical power. Also the correlation coefficients seen between hit-ratio and the MMN amplitude in the present study were only around 0.5, compared with much higher correlations in prior studies with simpler duration and frequency deviancies (Amenedo & Escera, 2000; Novitski et al., 2004). However, recent studies with more complex music and speech stimuli have demonstrated correlations more comparable to the ones seen in the present study (Earle, Landi, & Myers, 2017; Virtala, Huotilainen, Partanen, & Tervaniemi, 2014).

Our results showed that the size of the MMN in response to the vowel deviant in the attentive condition correlated positively with the hit-ratio of the vowel deviant in the post-training condition. In addition,

the size of the MMN to the rule violations in the attentive condition correlated positively with the hit-ratio of the rule violations in the training sequence. The associations obtained between the MMN response amplitudes in attentive condition and task performance, although rather small, are consistent and in line with previous findings (e.g., Atienza et al., 2002; Amenedo & Escera, 2000; Jaramillo et al., 2000; Novitski et al., 2004).

However, these associations were not significant for the MMNs in the ignore condition, for the vowel-hit-ratio in the training sequence, or for the rule-hit-ratio in the behavioral detection task. Also, no correlation was demonstrated between the MMN amplitude and behavioral reaction time, unlike in some previous studies (Amenedo & Escera, 2000; Tiitinen, May, Reinikainen, & Näätänen, 1994). For example Amenedo and Escera (2000) demonstrated that the MMN amplitude in response to duration changes in spectrally rich sounds, recorded in a preceding ignore condition, was highly correlated with reaction time *and* accuracy in a separate behavioral detection task following the MMN recording.

A number of other, less systematic correlations were seen in the present study between MMN and P3a properties and task performance.The MMN latency in an ignore listening condition and reaction time in a subsequent behavioral detection task were highly positively correlated in a previous study with frequency changes in sinusoidal sounds (Tiitinen et al., 1994). However, a weak negative correlation was seen in the present study: for vowel deviants, longer MMN latency during the ignore condition predicted shorter reaction times in the post-training condition. This unexpected result could be related to attention effects: the participants who performed accurately in vowel deviant detection, may have had difficulties in suppressing the salient vowel deviants in the ignore condition. Therefore, attention-related components overlapping and following the MMN (like the N2b), may have been particularly pronounced in those participants, resulting in the MMN peak latencies appearing later when task performance was fast and accurate. Also, it is important to note that the vowel deviants were naturally very well detected by the participants, resulting in a narrow range of hit-ratios with a very skewed distribution. This is likely to compromise the results obtained for the vowel deviant. Furthermore, in the present study, reaction time in the behavioral detection of rule violations was associated with the P3a amplitude in the attentive condition: longer reaction time was associated with *larger* P3a's, contrasting the prior opposite finding by Novitski et al. (2004).

These somewhat inconsistent and weak findings in the present study compared to previous work regarding correlations between ERPs and behavioral detection performance may be at least partly explained by the stimuli used. While previous studies have presented simple stimuli varying in some basic sound feature ranging from easy to hard to discriminate (Amenedo & Escera, 2000; Novitski et al., 2004; Tiitinen et al., 1994), the present study used complex speech stimuli with either highly salient (vowel) changes or very subtle changes (rule violations). Thus, previous findings are not directly comparable with the present results which were obtained in a novel, acoustically complex paradigm. Overall, the present results support the idea that the relationship between neural change-discrimination responses and perceptual, behavioral level of deviance detection is not straightforward, particularly in natural, complex listening conditions. In future studies, robust statistical methods and large sample sizes are essential in order to gain a proper understanding on the associations between MMN, P3a and task performance also in the case of natural sounds and complex discrimination tasks.

4.4. Implications for speech processing

Speech is an essential stimulus for the human auditory system. Tolerance for acoustic invariance, sensitivity to native language phoneme categories, and detection of complex sound relationships underlying, e.g., speech prosody and phonotactic rules, are central requirements for efficient speech processing. The present study offers a novel, ecologically valid paradigm to study speech processing in complex, natural conditions. By introducing acoustic variation to a stream of speech sounds, the paradigm we used mimics natural spoken language with, e.g., prosodic and speaker-related variability. The rule included in the auditory sequence serves as a regularity resembling those that exist in language, for example, the complex phonotactic rules of speech as well as the subtle acoustic changes that carry meaning in communication. Prosodic speech features, such as word or phrase boundary cues or paralinguistic cues used to denote, e.g., sarcasm (Rankin et al., 2009; Rockwell, 2007) or to emphasize specific parts of speech, are relevant examples of such subtle changes, highly important to master in order to perceive speech in a socially adequate, intelligent manner.

By shedding light on these phenomena in healthy adults, the present work offers a basis for future studies in clinical child and adult populations. There is already evidence that in developmental dyslexia and other language disorders, auditory paradigms including small deviances and complex sounds are more sensitive to reveal processing deficits than larger deviances embedded in simpler, e.g., sinusoidal sounds (Kujala, 2007). Furthermore, it was shown that in autism, different MMN results are obtained when speech sound changes are presented in simple sequences with repetitive sounds, compared to when they are presented in an acoustically variable (F0-variation) context (Lepistö et al., 2008). Evidently, too simplified paradigms might yield misleading information on the nature of neural processes underlying language and related disorders.

Specifically, our results on the automaticity of the rule violation detection, as evidenced by the MMN, suggest that the difficulties in detecting minute changes in language might be due to several different causes. For example, deficient discrimination ability might suggest impaired change-detection ability at the pre-attentive level, but it could also result from difficulties in translating this automatic change-detection to explicit awareness.

5. Conclusions

To conclude, the present study demonstrates that the human auditory system is capable of detecting phoneme changes and violations of abstract rules in an acoustically varying native language speech context. This detection occurs implicitly, at the pre-attentive processing level, as explicit awareness of the nature of the changes was not a prerequisite for their neural discrimination. However, our results suggest that different neural processes play different roles in discrimination and gaining explicit awareness of the changes: while the MMN is elicited regardless of whether one is explicitly aware of the changes, MMN amplitude is still associated with behavioral detection accuracy which, in turn, predicts the emerging explicit awareness of the heard changes. On the other hand, the P3a seems to work at the boundary of the explicit and the implicit awareness of the changes and thus play a different role in auditory discrimination. On the basis of the present results, we propose that the implicit neural discrimination of speech sound changes is likely to serve as a basis for efficient and fluent speech perception and language learning.

Acknowledgements

References

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I., et al.

(1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology, 35*, 211–224.

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience, 12*, 2570–2574.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning & Memory, 9*, 138–150.

Bendixen, A., & Schröger, E. (2008). Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biological Psychology, 78*, 231–241.

Bendixen, A., Roeber, U., & Schröger, E. (2007). Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience, 19*, 1664–1677.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer [Computer software].* [Retrieved from: www.praat.org].

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport, 12*, 3155–3158.

Earle, F. S., Landi, N., & Myers, E. B. (2017). Sleep duration predicts behavioral and neural differences in adult speech sound learning. *Neuroscience Letters, 636*, 77–82.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*, 244–249.

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology, 79*, 139–147.

Jaramillo, M., Paavilainen, P., & Näätänen, R. (2000). Mismatch negativity and behavioural discrimination in humans as a function of the magnitude of change in sound duration. *Neuroscience Letters, 290*, 101–104.

Kujala, T., & Näätänen, R. (2010). The adaptive brain: A neurophysiological perspective. *Progress in Neurobiology, 91*, 55–67.

Kujala, T. (2007). The role of early auditory discrimination deficits in language disorders. *Journal of Psychophysiology, 21*, 239–250.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological Psychology, 74*, 1–19.

Lang, S., & Kotchoubey, B. (2000). Learning effects on event-related brain potentials. *Neuroreport, 11*, 3327–3331.

Lepistö, T., Kajander, M., Vanhala, R., Alku, P., Huotilainen, M., Näätänen, R., et al. (2008). The perception of invariant speech features in children with autism. *Biological Psychology, 77*, 25–31.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology, 48*, 4–22.

Novitski, N., Tervaniemi, M., Huotilainen, M., & Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research, 20*, 26–36.

Paavilainen, P., Jaramillo, M., Näätänen, R., & Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters, 265*, 179–182.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology, 38*, 359–365.

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *Neuroreport, 18*, 159–163.

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology, 88*, 109–123.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia, 50*, 341–349.

Potts, G. F., Dien, J., Hartry-Speiser, A. L., McDougal, L. M., & Tucker, D. M. (1998). Dense sensor array topography of the event-related potential to task-relevant auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 106*, 444–456.

Rankin, K. P., Salazar, A., Gorno-Tempini, M. L., Sollberger, M., Wilson, S. M., Pavlic, D., et al. (2009). Detecting sarcasm from paralinguistic cues: Anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage, 47*, 2005–2015.

Ritter, W., Paavilainen, P., Lavikainen, J., Reinikainen, K., Alho, K., Sams, M., et al. (1992). Event-related potentials to repetition and change of auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 83*, 306–321.

Rockwell, P. (2007). Vocal features of conversational sarcasm: A comparison of methods. *Journal of Psycholinguistic Research, 36*, 361–369.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *Neuroreport, 3*, 1149–1151.

Schröger, E., Bendixen, A., Trujillo-Barreto, N. J., Roeber, U., Wallis, J., Anderson, K., et al. (2007). Processing of abstract rule violations in audition. *PLoS ONE, 2*, e1131.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *Neuroreport, 13*, 1813–1816.

Sussman, E. S. (2007). A new view on the MMN and attention debate. *Journal of Psychophysiology, 21*, 164–175.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*, 90–92.

Trask, R. L. (1996). *A dictionary of phonetics and phonology.* London: Routledge.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport, 9*, 3557–3560.

Virtala, P., Huotilainen, M., Partanen, E., & Tervaniemi, M. (2014). Musicianship facilitates the processing of Western music chords –An ERP and behavioral study. *Neuropsychologia, 61*, 247–258.

van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Näätänen, R., & Tervaniemi, M. (2006). Implicit, intuitive: and explicit knowledge of abstract regularities in a sound sequence: An event-related brain potential study. *Journal of Cognitive Neuroscience, 18*, 1292–1303.

Wiik, K. (1965). *Finnish and English vowels. A comparison with special reference to the learning problems met by native speakers of Finnish learning English.* Turku: University of Turku: Publications of University of Turku, B: 94.