Non-native phonemes in adult word learning: evidence from the N400m

Christian Dobel[1,3,*], Lothar Lagemann[1] and Pienie Zwitserlood[2,3]

[1]*Institute for Biomagnetism and Biosignal Analysis,* [2]*Department of Psychology, and*
[3]*Otto-Creutzfeldt-Centre for Cognitive and Behavioral Neuroscience, Westfälische Wilhelms-Universität Münster, 48149 Münster, Germany*

Newborns are equipped with a large phonemic inventory that becomes tuned to one's native language early in life. We review and add new data about how learning of a non-native phoneme can be accomplished in adults and how the efficiency of word learning can be assessed by neurophysiological measures. For this purpose, we studied the acquisition of the voiceless, bilabial fricative /Φ/ via a statistical-learning paradigm. Phonemes were embedded in minimal pairs of pseudowords, differing only with respect to the fricative (/aΦo/ versus /afo/). During learning, pseudowords were combined with pictures of objects with some combinations of pseudowords and pictures occurring more frequently than others. Behavioural data and the N400m component, as an index of lexical activation/semantic access, showed that participants had learned to associate the pseudowords with the pictures. However, they could not discriminate within the minimal pairs. Importantly, before learning, the novel words with the sound /Φ/ showed smaller N400 amplitudes than those with native phonemes, evidencing their non-word status. Learning abolished this difference indicating that /Φ/ had become integrated into the native category /f/, instead of establishing a novel category. Our data and review demonstrate that native phonemic categories are powerful attractors hampering the mastery of non-native contrasts.

Keywords: phoneme learning; N400; magnetoencephalography; statistical learning

1. INTRODUCTION

The current study investigates whether adults can acquire phonemes of a non-native language in an untutored way, and how words containing such speech sounds interact with already existing semantic representations. The adult participants in our study could learn by—more or less unconsciously—exploiting statistical probabilities, which also helps young infants to determine the specific acoustic properties of their native language. To determine the effectiveness of learning, we employ behavioural and neurophysiological measures such as the N400m. The theoretical and empirical underpinnings, as well as the rationale for our study, are based on insights from first-language acquisition research.

(a) *From words to phonemes and meaning*

Human language is a multi-layered system, governed by multiple principles and rules, from discourse and sentence level to the lower levels of, e.g. words, morphemes and phonemes. One critical question in psycholinguistics is how meaning is connected to the surface forms of spoken, written or signed language. While the discourse and sentence levels describe complex states, including interactions between several scene participants, single words such as nouns or verbs refer to more circumscribed states (e.g. a noun referring to a particular scene participant) and events (e.g. a verb referring to an action). But words are not the smallest unit of meaning. Words often consist of multiple morphemes, and even the bound ones are semantically interpretable, such as 'un-' in 'unchangeable', 'untouchable' or unbreakable'. Morphemes in turn consist of sequences of speech sounds, so-called phonemic segments or phonemes, such as /t/ as in 'cat', 'stand' or 'water'. Note that each realization of /t/ within these words sounds different: They are different phones. Phones are allophones within a particular language if they map onto the same phonemic category, despite of variation. Whether phones are allophones or different phonemes in a given language is tested by the existence of minimal pairs. A minimal pair consists of two word forms that differ in only one phoneme and, crucially, in meaning. In English, [dip] and [tip] are words with different meaning and, thus, /t/ and /d/ are different phonemes. In contrast, /t/ and /d/ are allophones in, for example, Korean, where [dip] and [tip] would be the same word. Thus, the existence of minimal pairs in a language proves that the two phones involved constitute different phoneme categories, and phonemes may thus be considered the smallest language units that can change the meaning of words. Given that phonemes often differ with respect to a single feature (such as manner or place of articulation, or voice), one could argue that features are atoms of speech.

The above example also demonstrates that languages differ with respect to their phoneme inventory. The size of the phoneme inventory ranges

* Author for correspondence (cdobel@uni-muenster.de).

between 10 and 80 phonemes. For example, German and English have 40–50 phonemes (depending on dialect). Comparing these values with other mental domains where a complex input has to be sorted into categories (such as face or object perception), this may not seem impressive. The challenge with speech is not in the number of categories but in the nature of the input. The physical realization of a phoneme depends heavily on speaker variability (due to differences in the articulatory apparatus, dialect and gender), speech rate and speech context (phonemes vary depending on adjacent speech). It still remains a mystery how infants—and adults—can cope with such variability in the input and still develop a highly effective speech comprehension system with remarkable speed.

Of crucial importance is that infants are well equipped from birth to deal with their native acoustic world, and that important language-relevant changes happen within the first year of life. Immediately after birth, newborns can discriminate the phoneme inventory of all languages, an ability that is no longer present in adults (Eimas *et al.* 1971; Eimas 1975; Lasky *et al.* 1975; Werker & Lalonde 1988; see Kuhl 2004, for a review). Thus, children come equipped with the potential to differentiate between phonemic contrasts of all languages. In addition to this potential, young children are able to categorize similar phonemic segments despite overt differences, due to different speakers, different speed or varying phonetic environments (Kuhl 1979, 2004; Miller & Liberman 1979; Eimas & Miller 1980; Hillenbrand 1984). In the course of their first year of life, babies lose the ability to distinguish all of the world's phonemes, as they become more and more tuned to the specifics of their mother tongue. Allophones become grouped together in language-specific phonemic categories (e.g. in Japanese /r/ and /l/ belong to the same category). While infants' abilities to discriminate and categorize native-language phonemic segments wax, their ability to discriminate non-native segments wanes (Cheour *et al.* 1998; Kuhl *et al.* 2001; Rivera-Gaxiola *et al.* 2005). Thus, as Janet Werker expressed it, children are born as 'citizens of the world' and become 'culture bound' listeners (Werker & Tees 1984). There is ample evidence that relearning non-native contrasts in adulthood is difficult, time-consuming and potentially requires explicit tutoring (cf. Menning *et al.* 2002).

(b) Statistical learning for language acquisition

An important mechanism responsible for the native-language tuning seems to be statistical learning (Saffran *et al.* 1996; Saffran 2002). It involves the implicit computation of the distributional frequency with which certain items (speech sounds, words) occur relative to others, or with which certain items co-occur. Speakers of different languages produce a wide range of similar sounds and their distributional frequencies can differ vastly across languages. Speech sounds that are typical for a particular phoneme category occur frequently, sounds at the border between categories are rare. Infants seem to compute distributional frequencies of sounds and categorize them according to their similarity to frequently occurring prototypes. In other words, native-language sounds display a 'perceptual magnet effect', that is, they attract similar sounds and facilitate categorization (Kuhl 1991; Kuhl *et al.* 1992; Sjerps & McQueen in press).

The mechanism of statistical learning not only applies to the acquisition of phonemes, but also to words. A main problem to grant lexical access is the discovery of word boundaries. Even though words are easily conceived of as unitary, distinct entities, spoken language appears most often without pauses for words. When we listen to foreign languages, we regularly have problems in isolating single words and the speech stream sounds like a connected string of sounds. However, a solution to this problem seems to be again the computation of sequential probabilities between connected syllables. Take as an example the noun phrase 'English artfulness'. 'Shart' is a non-existing word in English that could, however, according to phonetic and orthographic rules, well exist (i.e. it is a pseudoword). In contrast to this combination of syllables, (English) children will more often encounter the combination of syllables 'eng-' and '-lish', similarly as 'art-', '-ful-' and '-ness'. Thus, the computation of the sequential frequencies of the syllables helps children to establish word boundaries and determine from a continuous speech stream meaningful units like words (Goodsit *et al.* 1993; Saffran *et al.* 1996; Saffran 2002). Notably, this happens implicitly and with remarkable efficiency. Children under the age of two comprehend about 150 words and actively produce about 50 words (Fenson *et al.* 1993). From then on children acquire on average eight new words per day until adolescence in order to reach a mental lexicon containing 50 000 entries in a typical language.

In sum, newborns are equipped with the remarkable ability to differentiate speech-relevant acoustic features of any language. Within their first year of life, they seem to compute statistical distributions of properties in their native-language input. This leads to a fine-tuning of native phonemic categories, while at the same time children become less sensitive to phonemic differences of non-native languages.

Adults thus have to relearn non-native phonemic contrasts that they were perfectly sensitive to during their earliest time of life. We investigate this relearning in our current study, providing the opportunity of statistical learning as a strategy to acquire a novel, non-native contrast. We thus try to create an experimental situation that shares several important characteristics with an infant's situation. Our participants are confronted with novel spoken stimuli that—in due time—are paired frequently with a particular meaning (conveyed by pictures of an object). The statistical probability of this pairing arises over time, given that every to-be-learned novel word is paired equally often with the same object (the one to which that novel word should refer) as with a series of different objects. This mimics the situation of the child that hears a certain combination of sounds being repeated more often in the presence of a specific object compared with a range of other objects.

Participants have to decide intuitively if word and object belong together. Because no feedback is given, this is comparable to the situation of a child that just listens to language while taking in its environment.

This type of associative/statistical learning has been used successfully several times before, showing that a large novel vocabulary can be learned within a few days (Breitenstein & Knecht 2002; Knecht *et al.* 2004). Given the possibility that pictures and novel word combinations can be presented equally often for frequent (consistent pairing for those that enable learning) and infrequent pairings (random pairing for those that should not enable stable learning) during learning, this makes it a useful and powerful tool for neuroscientific investigations (Breitenstein *et al.* 2005). Moreover, this type of learning does neither involve complex instructions nor does it require a lot of attention. Computer implementations make intensive training and training at home easily possible. As such it is perfectly suited for special populations such as aphasics or children with impaired language development. In fact, the statistical learning principle was already successfully applied in aphasic patients who were trained in object naming (Schomacher *et al.* 2006). Similarly, we demonstrated recently that children with cochlear implants and reduced vocabulary are able to acquire novel verbs of their native language with only a few days of training (Glanemann *et al.* 2008). Even though the training sessions were carried out in a relative passive and receptive mode as described above, there was a solid transfer effect to action naming. Thus, there is increasing evidence that statistical learning affects language representations in a deeply rooted manner. In a crossmodal priming study (Breitenstein *et al.* 2007), we recently showed that novel words acquired via associative/statistical learning evoke similar priming effects as words from one's mother tongue. It seems that this type of learning implements new vocabulary deeply into the mental lexicon and the semantic system. Importantly, statistical learning also withstands the test of time, given that learning success is still evident weeks after training.

In contrast to earlier studies, the novel stimuli of the current study have a special characteristic: participants have to learn novel words that form minimal pairs. The crucial difference between the members of a pair concerns a native/non-native speech sound contrast. While one novel word of each pair contains the German phoneme /f/ (as in [vaf]), the other contains the unvoiced bilabial fricative /Φ/ (e.g. [vaΦ]) at the same position (note that none of the novel stimuli were German words). Unvoiced bilabial fricatives do not exist in German and English (although they are present, non-contrastively, in certain Scottish and Irish dialects; Strevens 1960). The sound is reminiscent of a soft blow of air through opened lips. But the unvoiced bilabial /Φ/ is an existing phoneme in several African languages. In Ewe, a language spoken by three to five million people in the area of Ghana and Togo (Ameka 2001), /Φ/ and /f/ are contrastive, given that they create minimal pairs at the word level. We thus confronted our learners with non-native speech sounds. Strictly speaking, stimuli containing non-native phonemes do not conform to German phonology and are thus non-words, but for the sake of brevity we refer to all stimuli as novel words. Crucially, our subjects were not given explicit feedback during learning in order to mimic the acquisition situation of young children. As such our enterprise is to assess whether relearning of phonemic contrasts is possible by an implicit learning strategy in adults.

(c) *The N400 as a tool to measure word learning*

We assessed learning success with behavioural and magnetoencephalographic (MEG) measures. The behavioural task, administered after learning, was to select from a set of four objects the one belonging to the learned novel word. With MEG, we implemented a priming design before and after learning, with existing German words and novel words as primes to target pictures. The dependent measure was the N400m, the neuromagnetic counterpart to the N400 from electro-encephalography (EEG). MEG is as time-sensitive as EEG, but it is a reference-free method. In comparison with EEG, neuromagnetic signals of a neuronal generator do not become distorted through surrounding brain tissue and they are less affected by high skull impedance. In addition, most MEG settings have more MEG sensors than EEG channels and so it offers more confidence in estimating the neural generators of specific components. As such it is sometimes taken as '... the best combination of spatial and temporal resolution of non-invasive methods in common use' (Van Petten & Luka 2006, p. 285). Note, however, that both methods are sensitive to generators with different orientations and therefore combined EEG/MEG is most probably a preferable approach.

Neurophysiological measurements using EEG or MEG, and particularly the N400 component, are by now widely used in neurolinguistics. The N400 is a large negative-going potential, peaking around 400 ms after stimulus onset. Its negativity is enhanced by semantic violations and semantic mismatch (cf. Kutas & Hillyard 1980). This was shown with sentence context and word pairs (e.g. Rugg 1985; see figure 1 for a typical example for an N400 design), with read, heard or signed verbal material (Kutas *et al.* 1987). But the N400 is also elicited by other meaningful stimuli, such as line drawings, photos and environmental sounds (Holcomb & McPherson 1994; Van Petten & Rheinfelder 1995; Ganis *et al.* 1996; Plante *et al.* 2000). Much less is known about the sensitivity of the N400 to lexical characteristics, such as word frequency (Van Petten & Kutas 1990, 1991; Van Petten 1993). Thus, as concluded by Van Petten & Luka (2006), 'the N400 amplitude is a general index of the ease or difficulty of retrieving stored conceptual knowledge associated with a word, which is dependent on both the stored representation itself, and the retrieval cues provided by the preceding context' (p. 281).

As we argued recently (Dobel *et al.* 2009), this property makes the N400 a valuable tool to investigate word learning. While, for example, mismatch
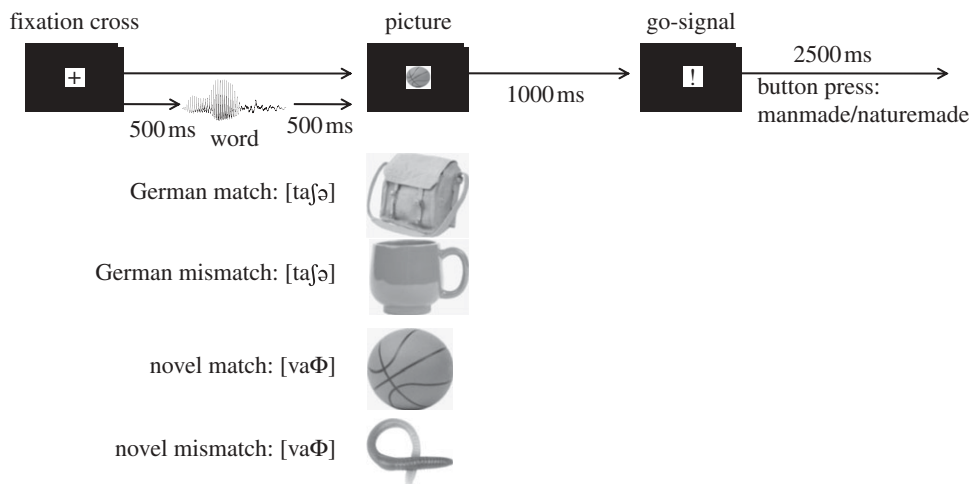
Figure 1. Time course (top) and conditions (bottom) of the crossmodal semantic priming procedure.

negativity can be used to investigate changes in auditory processing of (learned) stimuli (for a recent review see Pulvermüller & Shtyrov 2006), the N400 allows the measurement of how (novel) verbal stimuli interact with conceptual information, in our case conveyed by pictured objects.

Until recently, the N400 was only rarely employed to assess language learning. McLaughlin *et al.* (2004) investigated learning of French as a second language in a classroom setting. An attenuation of the N400, for French existing words compared with pseudowords, was observed after 14 h of instruction. An N400 reduction due to semantic relatedness became visible only after 60 h of teaching and was robust after about 140 h. That the N400 is sensitive to much faster learning was recently demonstrated by our own statistical word-learning study (Dobel *et al.* 2009) and by Mestres-Missé *et al.* (2007), who had participants infer the meaning of pseudowords embedded in meaningful sentence context. Amazingly, after only three exposures, participants' brain responses to such pseudowords were indistinguishable from real words. This was not the case for pseudowords for which no distinct meaning could be inferred. Nevertheless, one should keep in mind that typical classroom teaching, like in the study by McLaughlin and co-authors, targets not only mere word learning, but also syntax, phonology, conversational skills, etc.

Here, as in our earlier studies, we combine statistical learning, over a number of days, with behavioural and MEG measures. In a nutshell, our previous results showed that the ability to correctly judge whether a novel word and an object belong together almost reached ceiling after only 1 h of training. Training was distributed over three short sessions on consecutive days (Breitenstein *et al.* 2007; Dobel *et al.* 2009). Relative to the pre-test, brain responses to objects primed by learned novel words (that were pseudowords before training) were strongly attenuated and not differentiable from responses to the same pictures primed by real words. Source localization showed that the left temporal cortex was the crucial structure involved in this change of activity from pre- to post-learning.

In the present study, we go beyond 'mere' word learning and ask whether adults can learn novel words when they come in pairs whose members only differ with respect to the native/non-native fricative. The critical test for successful learning of non-native phonemes is the correct discrimination of the members of minimal pairs. We tested this with behavioural means, and with crossmodal priming, assessing the impact, in terms of N400m amplitude, of each novel word on the object with which it was associatively paired, as well as on the object that was paired with its direct companion minimal word. During the MEG recording, participants had to decide if the pictured objects are man-made or natural. To assess the size of strong and weak context effect from the native language of the participants, we also included spoken German words in related and unrelated conditions in the MEG experiment. These words also came in minimal pairs (e.g. /bus/, bus and /bus/, bush; /glas/, glass and /gras/, grass). A minimal native-language phonemic difference should prevent priming of the meaning of the companion word. During learning, the pictures for these German word pairs were combined with novel words (e.g. picture of a bus paired with /afa/, picture of a glass with /aΦa/). We use source localization methods to establish the crucial neural structures of such potential learning.

2. MATERIAL AND METHODS
(a) *Participants*
Fifteen participants (mean age: 26 years; nine women; all right-handed native speakers of German with comparable level of education) took part in the experiment. They all had normal or corrected-to-normal vision. None had participated in any of the earlier learning studies.

(b) *Auditory stimulus material*
We first selected 24 German minimal pairs, one or two syllables long, from the CELEX corpus (Baayen *et al.* 1995). The main criterion was whether their meaning could be visualized easily. Half referred to man-made objects, the other half to nature-made objects.

The differing phonemes within a minimal pair were always consonants, equally distributed over initial, medial or final position within the words. The 48 pictures belonging to the German words were randomly paired with the novel words of the 24 minimal pairs (randomization was done with the EXCEL 2003 software (Microsoft Corporation) by inserting random numbers between 0 and 1 next to the list of words and sorting them in ascending order; then this randomly sorted list of words was merged with the picture list of equal length). Within a pair, the two novel stimuli differed only in the phoneme /f/ or /Φ/, again distributed equally over word-initial, -medial or -final position. All pseudowords were mono- or disyllabic (e.g. /afa/ and /aΦa/; /vaf/ and /vaΦ/). Except for /Φ/, all phonemes were part of the German inventory.

All stimuli were spoken by three trained male speakers in a recording studio at the university's radio station. This was done to have three different exemplars for each word. For recording we used a 'Hitec Audio Fat One' condenser microphone and a 'Studer Typ 962' mixing desk. Sounds were digitized with a 'Creamware Pulsar2' sound card with a sampling rate of 44 100 Hz and stored as individual sound files in wav-format using the software CUTMASTER. The stimuli were excised, at as close as possible to their on- and offsets, with the software package PRAAT and scaled to an equal RMS amplitude of -24 dBFS.

During MEG recording, SRM-212 electrostatic earphones (Stax, Saitama, Japan) were used to transduce air-conducted sounds. All sounds were delivered through silicon tubes (length: 60 cm; inner diameter: 5 mm) and silicon earpieces, which were adjusted to fit to each subject's ears. Because auditory stimuli were delivered via a tube system into the MEG chamber, the audibility of items was measured with two judges. They heard all pseudowords via the tube system and had to decide via forced choice if the word contained /f/ or /Φ/. Response accuracy was at 100 per cent in both cases. Note that in contrast to the experimental participants these judges were fully informed about the phonemic differences in the pseudowords. As such this test only made sure that the differences were in fact audible.

(c) Visual stimulus material
A total of 192 colour photos of objects were used. Each of the 48 objects (for 24 minimal pairs, each with two words) was presented in four different variants. These objects represented the meaning of the 48 words of the 24 German minimal pairs. The pictured objects were randomly combined with novel words during learning (see explanation above for the randomization procedure). Note that different pictures of the same objects were used during training. All pictures stem from Hemera Photo-Objects Vol. I–III or from Internet sites (http://flickr.com; http://commons.wikimedia.org). Using IrfanView 3.98, all images were resized to 7 cm width (72 dpi). All 192 pictures were rated by nine participants. They had to name the pictures and to judge, on a scale from 1 to 5 ('1 = fits very well'; '5 = does not fit at all'), whether the intended names fitted to the pictures. All selected pictures were named with the intended name or a synonym and all were minimally judged as 'fitting well'.

(d) Novel word training and behavioural measures during and after learning
For the training with the novel vocabulary, the object pictures of the German minimal pairs (e.g. bus and bush; glass and grass) were combined with novel pseudowords. The two pictures that belonged to a minimal word pair in German were distributed over different minimal pairs in the novel vocabulary (e.g. bus pictures paired with /afa/, glass pictures with /aΦa/).

Participants were trained with the new vocabulary for about 48 min d^{-1}, on 5 consecutive days (Mondays through Fridays). In the course of 5 days, the 48 novel words were presented 80 times (i.e. eight times per day), 40 times with the same object, but only four times with each of 10 different objects. This procedure ensured that all novel word forms and all objects appeared equally often during training. On each learning trial, objects were presented 200 ms after the onset of the auditory stimulus presentation. Objects remained on the screen for 1500 ms, during which participants had to indicate via forced-choice button press whether pseudowords and objects matched. They were encouraged to do this spontaneously, on an intuitive basis. There was no feedback, and no information was given about the underlying learning principle. Participants learned according to one of two lists, with different novel word–object pairings, to avoid chance associations between specific stimuli. On the first day, they were told about the presence of the novel phoneme and the existence of minimal pairs in the novel vocabulary.

After the last day of learning, participants performed a multiple-choice test. On each trial, four pictures were presented on the screen. Two pictures were associated with the words from a novel minimal pair; the other two were from the remaining set. One of the novel words was presented 500 ms before each presentation. Participants had to decide via key press which of the four pictures corresponded to the heard novel word. The keys corresponded to the visual arrangement of pictures (e.g. top left picture corresponding to top left key).

Learning was performed with the auditory tokens of two speakers, while tokens of the third speaker were used for MEG recording only.

(e) MEG crossmodal priming experiment: pre- and post-training
In the MEG crossmodal priming paradigm administered before and after training, each of the 48 objects was presented four times, preceded by four different auditory primes: (1) the correct German picture name (German match), (2) the corresponding German minimal-pair partner (German mismatch), (3) the novel word that had been paired with the picture during training (novel match), and (4) its minimal-pair partner (novel mismatch). For an illustration see figure 1. Note that conditions (2) and (4) mismatch the picture before and after learning.

In contrast, condition (1) is always a match condition, while condition (3) is a mismatch condition before learning and a potential match condition after learning.

Primes were presented auditorily. After an interstimulus interval of 500 ms, an object picture was displayed for 1000 ms. To avoid preparatory movement artefacts, there was a response delay. An exclamation mark presented 2500 ms after picture object offset signalled that participants should indicate via button press whether the object was man-made or nature-made. Thus, decisions were explicitly delayed, which makes the latency data uninteresting. After the button press, a fixation cross was presented and remained on the screen until presentation of the next object.

The four prime conditions of each object were presented in random order and care was taken that not more than three trials requiring a 'correct' response appeared in sequence (the randomization procedure was done similarly as above, but after generating a random order the sequence was checked and altered manually if necessary). Similarly, minimal pairs were separated by at least five trials. A recording session lasted about 15 min.

(f) MEG recording and analysis

MEG signals were recorded by means of 275-sensors whole-head MEG system (Omega 275, CTF, VSM MedTech Ltd.) with first-order axial SQUID gradiometers (2 cm diameter, 5 cm baseline, 2.2 cm average inter-sensor spacing). Data were recorded continuously with first-order gradient filtering at 275 sensors. Responses were sampled at 600 Hz and filtered online with a frequency band-pass of 0–150 Hz. Recordings were further processed offline using Brain Electrical Source Analysis (BESA). Data were filtered using a 48 Hz low-pass and a 0.01 Hz high-pass filter. The averaging epoch was defined from 300 ms before picture onset to 700 ms after (-300 to 700 ms) with a 150 ms prestimulus baseline. Trials exceeding 3 pT were excluded from averaging.

To evaluate the underlying neural activity, source-space activity was estimated for each time point in each condition and subject, using the least square minimum-norm estimation (L2-MNE) method (Hämäläinen & Ilmoniemi 1994). This inverse source modelling and the consecutive statistical analysis were conducted with the MATLAB-based EMEGS software (www.emegs.org). The L2-MNE is an inverse method allowing the reconstruction of distributed neural sources underlying the extracranially recorded event-related magnetic fields, without the necessity of *a priori* assumptions regarding the number and possible locations of underlying neural generators. The L2-MNE is calculated by multiplying the pseudoinverse of the so-called lead-field matrix (which describes the sensitivity of each sensor to the sources) with the averaged recorded data. Individual lead-field matrices were computed for each participant, based on information about the centre and radius of a sphere fitting best to the digitized head shape and the positions of the MEG sensors relative to the

head. A spherical shell with 8 cm radius and with 350 evenly distributed dipole locations served as distributed source model. At each dipole location, two perpendicular dipoles were positioned, which were tangentially oriented to the spherical model.

The results of the L2-MNE solution are source wave forms over time for each dipole location (vector length of the corresponding tangential dipoles). Visual inspection of the Global Power (mean squared activity across all sources and time points) of the Grand Mean of the L2-MNE solutions for the German mismatch condition in comparison to the German match condition was used to establish the time intervals for the N400m, ranging between 350 and 500 ms, with a broad peak between 400 and 500 ms. The selection of this interval was confirmed by a point-wise repeated-measurement analysis of variance (ANOVA) for each time point and dipole location, with the factor MATCH (German match versus German mismatch). The Global Power of F-values for this factor displayed continuously significant ($p < 0.01$) F-values for this interval (i.e. more than 90 sequential data points) between 350 and 500 ms. The statistical parametric F-values of the main effect of MATCH were projected on a standard cortical surface (MNI (Montreal Neurological Institute) brain contained in EMEGS), in order to display the origin of effects in more detail. Because modelling was done with a sphere without realistic head and volume conductor models, it is not possible to report exact coordinates of the effects.

The values from the data-driven derived regions of interests were used for statistical analyses, with a repeated-measurement ANOVA with the factors SESSION (pre- versus post-learning), LANGUAGE (German versus novel vocabulary), MATCH (match versus mismatch) and HEMISPHERE (left versus right). Two-sided paired t-tests were used for post hoc comparisons. Note that the point-wise ANOVA is likely to produce significant or high-significant results by chance. In order to avoid such misinterpretations, we rely after this analysis only on long-lasting and robust effects. An obvious disadvantage of this procedure consists in the potential overlooking of small, short-lived effects. We consider as the major advantage the completely data-driven analysis of highly complex data.

To assess the pre- and post-changes of novel words as a function of whether they contained native or non-native phonemes, we separated these stimuli that had been averaged together in the novel mismatch condition (4). We did this for the mismatch condition, because we expected the strongest signal, which is necessary for source localization, since the reduction of trials for averaging affects the signal-to-noise ratio. We performed again a point-wise ANOVA for each time point and source, with the factors SESSION (pre- versus post-learning) and PHONEME (native versus non-native). Mapping the F-values of the interaction SESSION × PHONEME on a cortical surface served to establish the crucial region of interest for this interaction. This region was used for further comparisons with a repeated-measurement ANOVA with the factors SESSION (pre- versus post-learning) and

PHONEME (native versus non-native). This analysis of multi-channel recordings (EEG and MEG) has become an established procedure for sensor and source space (for recent studies see e.g. Kissler *et al.* 2007; Schupp *et al.* 2007; Dobel *et al.* 2008).

The study conforms with The Code of Ethics of the World Medical Association (Declaration of Helsinki). It falls under the ethical approval of the 'Kommission der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen-Wilhelms Universität Münster'.

All data of the study will be stored for 10 years and are available via the corresponding author.

3. RESULTS

(a) Behavioural data

During the course of training, correct responses (hits and correct rejections) progressed linearly from slightly above chance level (58%) at the end of day 1 to 88 per cent on day 5 ($F_{(4,56)} = 65.35$; $p < 0.001$). Planned contrasts suggest a linear effect ($F_{(1,14)} = 125$; $p < 0.001$). There was a significant increase of correct responses from day 1 to day 4 (all $t > 3.5$; all $p < 0.01$). A significant, but much smaller quadratic effect ($F_{(1,14)} = 13.5$; $p = 0.002$) was also visible due to the asymptotic slope of typical learning curves. Participants did not further improve from day 4 to day 5 ($t = 0.9$; n.s.). Learning happened very rapidly, as evidenced by a correct mean response rate of 81 per cent on day 3.

The multiple-choice test demonstrated that participants chose the correct object of the learned word in 49 per cent of cases; in 45 per cent of the cases they selected the object of the corresponding minimal-pair partner ($t_{(14)} = 0.9$; n.s.). Clearly, they failed to distinguish between the two pseudowords of a minimal pair, associating each pseudoword equally to both objects. In contrast, they only chose a wrong object, unrelated to the minimal pair, in 6 per cent of the cases.

(b) Magnetoencephalographic data

Inspection of the Global Power L2-MNE values (see figure 2, plotted for the relevant time duration (-100 to 550 ms) with $t = 0$ corresponding to the presentation of the object) demonstrated that prior to learning (top), the German mismatch condition evoked greater activity than the German match condition.

This activity was quite sustained, starting around 300 ms and lasting at least until 500 ms after picture onset. Based on visual inspection, information on the common N400 interval from the literature and the point-wise ANOVA, we determined the interval between 350 and 500 ms as the interval of interest for the N400 m. Figure 2a also shows that novel words induced stronger brain activity than both German conditions (for better visualization, the two novel-word conditions were collapsed in figure 2a). After learning, the German mismatch condition again evoked greater activity than the German match condition. The activity of both novel-word conditions (match and mismatch), however, was strongly
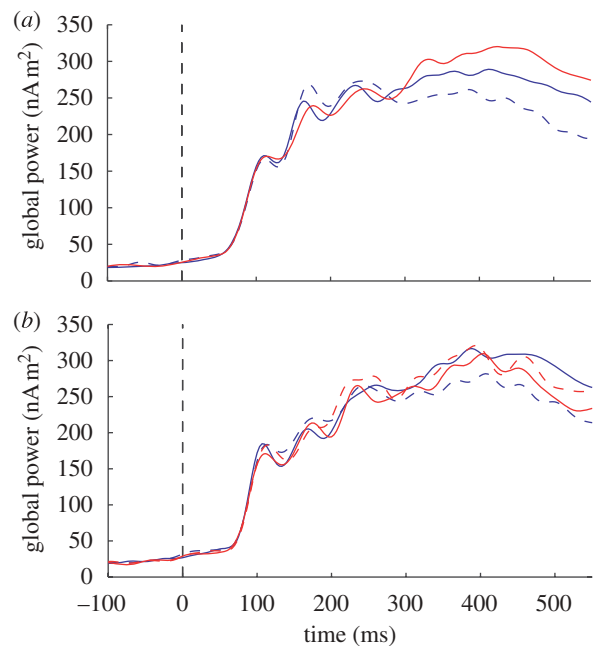

Figure 2. Global-power plot across all dipoles for all conditions (*a*) before and (*b*) after training. Blue, German; solid, mismatch; dashed, match; red: novel. For the visual display, data were low-pass filtered with 30 Hz.

reduced, and now located between the two German conditions (figure 2b).

Projecting the *F*-values of the factor MATCH for the German match and mismatch conditions indicated that a broad region of the left temporal lobe and the temporoparietal junction was crucially involved in differentiating these German conditions (figure 3, top). Figure 3 (bottom left) displays the time course of activity for this region only (plotted for the relevant time duration (-100 to 550 ms) with $t = 0$ corresponding to the presentation of the object). Note that there seem to be two active regions in figure 3 (top), one inferior and one superior in the temporal cortex. We included this as a factor in the described ANOVA. There was a main effect of this factor with greater activity for the superior compared with the inferior region, but no significant interactions with the factors SESSION, LANGUAGE and MATCH. Thus, inclusion of this factor would not change the results qualitatively. Note there seemed to be also a smaller effect in the frontal inferior gyrus. Because this effect turned out to be short lived in comparison with the remaining effects, we do not report it in more detail.

Based on this region of interest, the repeated-measures ANOVA with the factors SESSION, LANGUAGE and MATCH confirmed what was visible in the global power (figure 3). There was a main effect of LANGUAGE ($F_{(1,14)} = 11.39$; $p = 0.005$) which was due to more activity in the novel-word conditions than in the German conditions. As expected, activity was increased in response to the mismatch as compared with the match conditions (main effect MATCH: $F_{(1,14)} = 10.20$; $p = 0.006$). These main effects were modulated by two interactions. While the novel vocabulary evoked
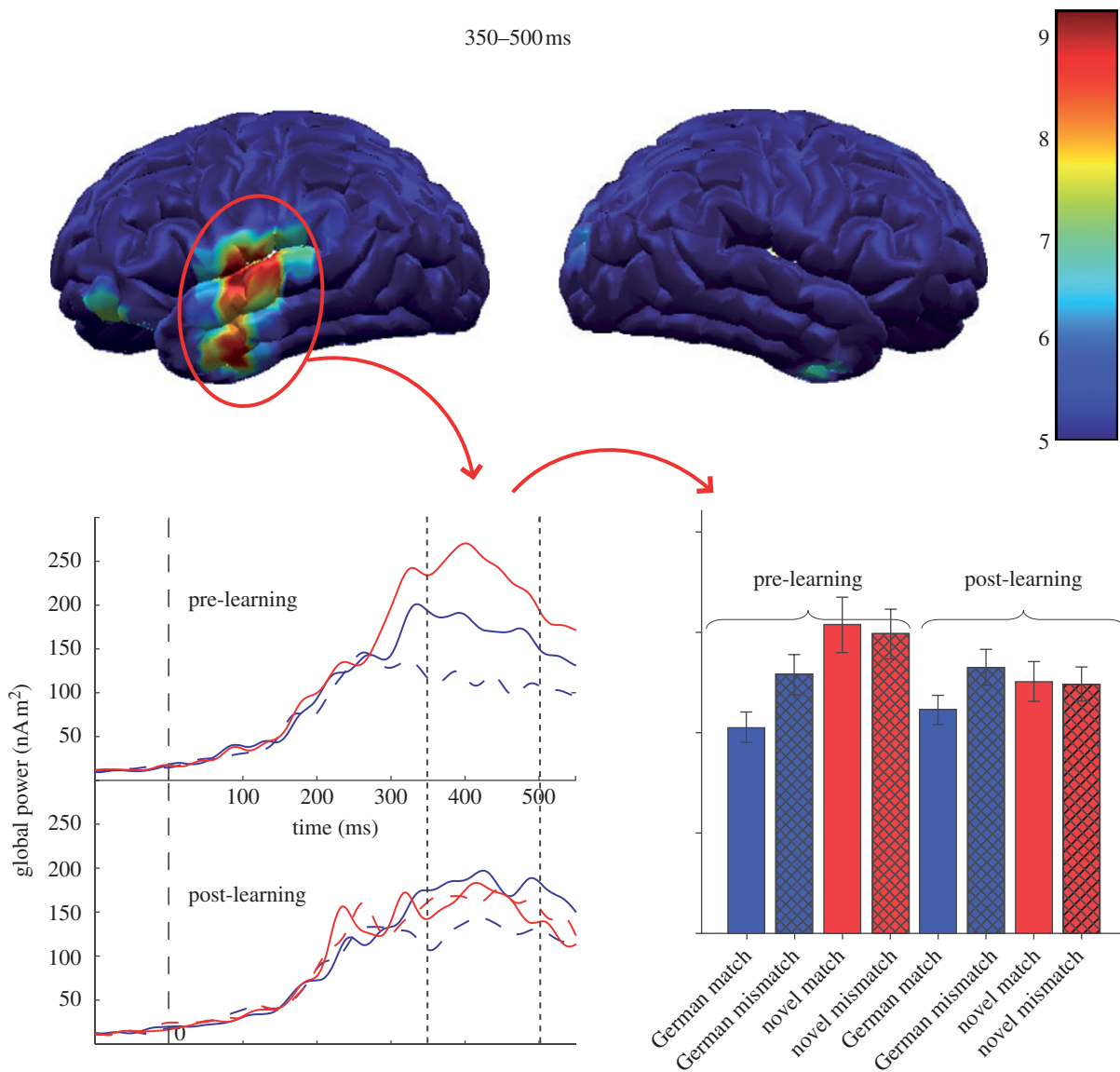
Figure 3. Top of figure displays the statistical parametric *F*-values of the effect MATCH for German, mapped on a standard cortical surface and averaged between 350 and 500 ms. The difference between matches and mismatches can be clearly seen over the left temporal gyrus. Left bottom of figure displays the time course of activity for this region (for the visual display data were low-pass filtered with 30 Hz). Right bottom figure shows the MNLS values over the left temporal area, broken down by session, language and match (left: pre-training; right: post-training; blue: German; solid: match; hatched: mismatch; red: novel; error bars denote standard errors).

greater overall activity during picture processing than the German vocabulary before learning ($t(14) = 4.42$; $p = 0.001$), this was not the case after learning ($t(14) = 0.56$; n.s.). These effects produced the interaction LANGUAGE × SESSION ($F(1,14) = 18.01$; $p = 0.001$). The interaction LANGUAGE × MATCH ($F(1,14) = 17.97$; $p = 0.001$) resulted from a strong N400m mismatch effect for German (mismatch versus match: $t(14) = 4.73$; $p < 0.001$), which was not present for the novel vocabulary (mismatch versus match: $t(14) = 0.71$; n.s.). No other significant main effects or interactions were found (see figure 3 bottom right).

The split of the mismatch condition of the novel words into the native phoneme and the non-native phoneme stimuli allowed us to track the change of impact of these words on subsequent semantic processing. Mapping the *F*-values of the interaction SESSION × PHONEME on a cortical surface revealed a somewhat smaller, more anterior region of the left temporal lobe than in figure 3 as the crucial region of interest (see figure 4, top). This region showed a sustained significant effect ($p < 0.01$) between 350 and 450 ms after stimulus onset. The repeated-measurement ANOVA with the factors SESSION and PHONEME based on this region confirmed the significant interaction of the two factors ($F(1,14) = 32.56$; $p < 0.001$). Importantly, the non-native phoneme condition evoked a smaller N400m than the native phoneme stimuli before learning ($t(14) = 3.40$; $p = 0.004$), but not after ($t(14) = 1.3$; n.s.; see figure 4 bottom).
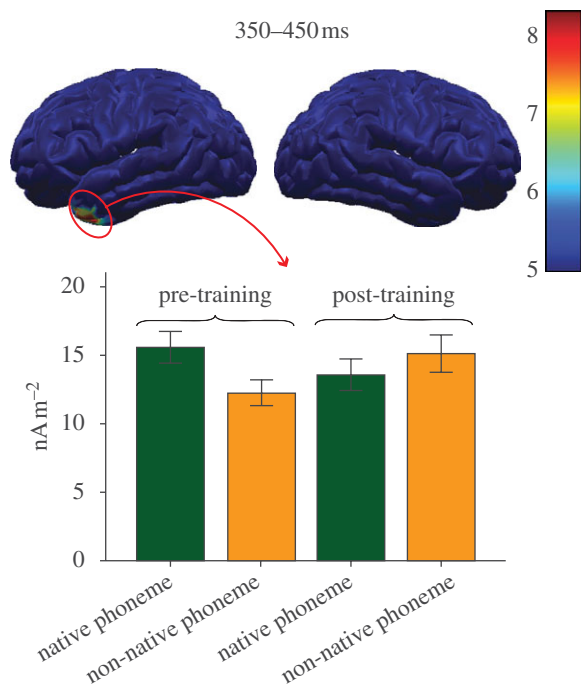
Figure 4. Top of figure displays the statistical parametric *F*-values of the effect SESSION × PHONEME mapped on a standard cortical surface and averaged between 350 and 450 ms. Bottom figure shows the MNLS values over the left temporal area, broken down by session and phoneme (left: pre-training; right: post-training; green: native, yellow: non-native; error bars denote standard errors).

4. DISCUSSION

We set out to investigate the acquisition of a non-native phoneme category in adults. The phoneme involved was the voiceless, bilabial fricative /Φ/, non-existent in German, embedded in 24 to-be-learned novel spoken stimuli, such as /vaΦ/ or /aΦo/. Each novel word had a minimal-pair partner, in which the non-native phoneme was replaced by a native fricative /f/ (e.g. /vaf/, /afo/). These 48 novel stimuli were trained with an untutored associative/statistical learning procedure, in which some combinations of novel words and pictures occurred more frequently than others. The goal was to establish a novel phoneme category, by implementing a contrast between a native and a non-native phoneme differing in only feature. Learning was evaluated by behavioural means, and by a crossmodal priming study, in which we measured the N400m for pictures preceded by existing German words and novel words.

As in earlier studies, the learning curve clearly demonstrated that the novel words were quickly associated with objects. After about only 2.5 h of training (at the end of day 3), response accuracy increased from chance level to about 80 per cent. A multiple-choice test after 5 days of learning showed that participants could almost perfectly decide which two objects belonged to a minimal pair of novel words. But they could not decide which of the two objects belonged to which member of the minimal pair. They thus were unable to differentiate between the /f/ and /Φ/ variant. Most probably, the non-native phoneme became integrated into the native category, as an allophonic variant. As a consequence, the distinction between novel words within a minimal pair was collapsed, and participants had learned to associate two meanings to each of these novel homonyms.

The N400m response corroborated this pattern of results. First, and as a test of the method, we obtained a native-language mismatch effect, where German related and unrelated conditions also only differed by one feature. As expected, the German mismatch words (/gras/ to a picture of a glass) induced a stronger N400m than matching words (e.g. /glas/—glass). Before learning, the novel words displayed an even stronger N400m than the unrelated German words. After learning, the N400m responses to the novel words were reduced and located between the German match and mismatch conditions. Interestingly, the N400m was indistinguishable for correct and incorrect pairings of novel words and objects within a minimal pair. This is completely in line with the behavioural data and corroborates our interpretation that two meanings were learned for each novel word, with its word form collapsing across the two allophonic variants.

If this interpretation is correct, the non-native phoneme should lose its foreign aspect during learning and become integrated as an allophone into the native category. Indeed, the N400m response was clearly sensitive to a change in the representational status of the non-native phoneme. Prior to learning, the N400m brain response in relation to non-native novel words was smaller than to the novel words with only native phonemes. This difference was absent after learning. Obviously, the novel phoneme had lost its non-native status during the course of learning. Below, we discuss the implications of these findings, starting with a discussion of the neural structure underlying the observed effects.

(a) *The N400 and non-native stimuli*

As in our recent study (Dobel *et al.* 2009), the crucial component for generation of the N400m was the left temporal lobe encompassing parts of the temporoparietal junction, as evidenced by distributed source modelling. Several studies with different methods collectively show that, depending on design and material, this structure is an important contributor to the N400 scalp potential (see Van Petten & Luka 2006, for a review). Moreover, patients with lesions in the left temporal lobe or temporal–parietal junction display strongly reduced and delayed N400 amplitudes (Hagoort *et al.* 1996; Swaab *et al.* 1997; Friederici *et al.* 1998). Intracranial recordings, probably the most direct method for the assessment of neuronal activity in a specific region, revealed activity in the anterior part of the left temporal lobe with very much the same time course and response to experimental manipulations as reported for the N400 (Smith *et al.* 1986; Nobre *et al.* 1994; Nobre & McCarthy 1995; Guillem *et al.* 1996; Elger *et al.* 1997; Fernández *et al.* 2001). Finally, several researchers before us have localized neuronal activity of the N400 in MEG. Whether by means of single and multiple dipoles, or with distributed sources, there is ample support that the temporal lobe and the

temporoparietal junction are involved in generating the N400 signal (cf. Simos *et al.* 1997; Helenius *et al.* 2002; Kwon *et al.* 2005; Maess *et al.* 2006). Of course, depending on stimulus material and tasks, other regions were similarly involved in generating the N400. For example, Maess *et al.* (2006) reported the additional involvement of left frontal sources using a similar methodology as we did. We assume that the activity of frontal areas is best explained by the employment of sentence material in their study.

Besides localization issues, our data add to a growing body of literature concerning the sensitivity of the N400m to particular stimuli and effects. The first concerns differences between words, pseudowords and non-words. We observed that novel words containing the non-native phoneme induced a weaker N400 response on the target pictures than novel words consisting only of native phonemes. This was the case before learning, but not after learning. This may sound surprising at first glance, but it fits well with what is known about the N400 in response to true non-words that violate native-language phonology in one way or another. Holcomb & Neville (1990) reported smaller N400 responses to non-words compared with phonotactically legal pseudowords. This held for the visual and auditory domain, with strings of consonants as visual non-words and words played backwards as auditory non-words. Similarly, Ziegler *et al.* (1997) found a clear reduction in N400 responses for illegal strings of consonants compared with pseudowords.

In line with these studies, prior to learning the N400m response was weaker when pseudowords contained a non-native phoneme, relative to pseudowords with only native phonemes. The combination of a non-native phoneme with native ones results in a string that is illegal according to German phonology and thus constitutes a non-word. What is truly interesting is that the stimuli containing non-native phonemes lost their non-word status during the course of learning. This change over time strengthens our conclusion that participants actually learned homonyms of a novel language, in which the sounds /f/ and /Φ/ exist and function as allophones. As such, our results corroborate the conclusions by Sebastian-Gallés *et al.* (2006) that the language system shows a remarkable lack to acquire non-native phonemic contrasts, but that first-language representations 'seem to be more dynamic in their capacity of adapting and incorporating new information' (p. 1277).

The second issue concerns materials and modalities. The reduced N400 component in our study was evoked by pictures, preceded by novel words containing native or non-native or speech sounds. The above studies measured the N400 on the non-words themselves (Holcomb & Neville 1990; Ziegler *et al.* 1997). The similarity of findings confirms that the N400 amplitude is an index of access to lexical and semantic representations for quite distinct stimuli in different modalities.

(b) *Learning novel phonemes*
The complete lack of a differentiation within the minimal pairs after learning, evidenced both behaviourally

and by the N400m response, clearly demonstrates that participants did not learn a novel phoneme. Instead, they categorized /Φ/ and /f/ as allophones. Thus, what we intended to be minimal pairs rather became homonyms after learning, like English 'bank' or 'bug'. We find it amazing that our learners were able to acquire such a large number of novel homonyms within such a short time, and with almost perfect accuracy. Given that (1) minimal pairs were easily distinguishable (as the pre-test had shown), (2) there were different word–object pairings between participants, (3) different visual and auditory tokens were used for learning and testing, and (4) that pseudowords showed different pre- and post-effects, we are confident that the effects were not due to stimulus artefacts.

Our findings seem to suggest that adults cannot acquire novel phonemes via associative/statistical learning, but we cannot be sure whether this applies across the board. First, we tried to establish a novel phonemic category that is in minimal contrast to an existing native one, differing in only one feature. Larger differences between native and non-native sounds, as, for example, with click consonants from the African language Xhosa, might have led to very different results. Next, we chose to follow established procedures in our training protocol (cf. Breitenstein & Knecht 2002; Breitenstein *et al.* 2007) and did not include feedback. Feedback might have improved the differentiation within minimal pairs. These arguments notwithstanding, we believe that the active production of novel phonemes is necessary for the establishment of a novel phonemic category, and we address this in further studies. Producing novel phonemes necessitates a focus of attention on differences between new and existing phonemes, which may well be a prerequisite for the robust implementation of a novel speech sound. A similar prediction follows from the motor theory of speech (Liberman *et al.* 1967; Galantucci *et al.* 2006).

Finally, we should keep in mind that 50 min of training per day might not be enough to change a deeply rooted representational system. Menning *et al.* (2002) trained German speakers to learn a moraic contrast of Japanese, involving durational differences, that occurs in many minimal pairs in Japanese (e.g. kite—kitte). Menning and colleagues trained German speakers with attention focused on the particular contrast, for $1.5\,\mathrm{h}\,\mathrm{d}^{-1}$, over the course of 10 days, with only four of such Japanese minimal pairs. After this extensive training, hit rate to differentiate the critical duration of morae increased by 35 per cent. Similar findings were reported by Bradlow *et al.* (1997).

Thus, a novel contrast within the native system that has been fine-tuned for more than 20 years seems to necessitate a more active and attentive training than statistical learning. Relearning contrasts that have been lost during early childhood is notoriously difficult and effortful. Our data document that native phonemic categories are powerful attractors in that they absorb the non-native stimulus, which is a considerable stumbling block on the path to the mastery of non-native contrasts.

REFERENCES

Ameka, F. K. 2001 Ewe. In *Facts about the world's languages: an encyclopedia of the world's major languages, past and present* (eds J. Garry & C. Rubino), pp. 207–213. New York, NY: H.W. Wilson.

Baayen, R. H., Piepenbrock, R. & Gulikers, L. 1995 *The CELEX lexical database (release 2).* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. & Tohkua, Y. 1997 Training Japanese listeners to identify English /r/ and /l/: some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* **101**, 2299–2310. (doi:10.1121/1.418276)

Breitenstein, C. & Knecht, S. 2002 Development and validation of a language learning model for behavioral and functional-imaging studies. *J. Neurosci. Methods* **2**, 173–180. (doi:10.1016/S0165-0270(01)00525-8)

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T. & Knecht, S. 2005 Hippo-campus activity differentiates good from poor learners of a novel lexicon. *Neuroimage* **15**, 958–968. (doi:10.1016/j.neuroimage.2004.12.019)

Breitenstein, C., Zwitserlood, P., de Fries, M., Feldhuis, C., Knecht, S. & Dobel, C. 2007 Five days versus a lifetime: intense associative vocabulary training generates lexically integrated words. *Restor. Neurol. Neurosci.* **25**, 493–500.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K. & Naatanen, R. 1998 Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* **5**, 351–353. (doi:10.1038/1561)

Dobel, C., Putsche, C., Zwitserlood, P. & Junghöfer, M. 2008 Early left-hemispheric dysfunction of face processing in congenital prosopagnosia: an MEG study. *PLoS ONE* **3**, e2326. (doi:10.1371/journal.pone.0002326)

Dobel, C., Junghöfer, M., Klauke, B., Breitenstein, C., Pantev, C., Knecht, S. & Zwitserlood, P. 2009 New names for known things: on the association of novel word forms with existing semantic information. *J. Cogn. Neurosci.* (doi:10.1162/jocn.2009.21297)

Eimas, P. D. 1975 Distinctive feature codes in the short-term memory of children. *J. Exp. Child Psychol.* **2**, 241–251. (doi:10.1016/0022-0965(75)90088-0)

Eimas, P. D. & Miller, J. L. 1980 Contextual effects in infant speech perception. *Science* **4461**, 1140–1141. (doi:10.1126/science.7403875)

Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. 1971 Speech perception in infants. *Science* **171**, 303–306. (doi:10.1126/science.171.3968.303)

Elger, C. E., Grunwald, T., Lehnhertz, K., Kutas, M., Helmstaedter, C., Brockhaus, A., Van Roost, D. & Heinze, H. J. 1997 Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia* **35**, 657–667. (doi:10.1016/S0028-3932(96)00110-8)

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. 1993 *The MacArthur communicative development inventories: user's guide and technical manual.* San Diego, CA: Singular Publishing Group.

Fernández, G., Heitkemper, P., Grunwald, T., Van Roost, D., Urbach, H., Pezer, N., Lehnertz, K. & Elger, C. E. 2001 Inferior temporal stream for word processing with integrated mnemonic function. *Hum. Brain. Mapp.* **14**, 251–260. (doi:10.1002/hbm.1057)

Friederici, A. D., Hahne, A. & von Cramon, D. Y. 1998 First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: electrophysiological evidence for a double dissociation. *Brain Lang.* **62**, 311–341. (doi:10.1006/brln.1997.1906)

Galantucci, B., Fowler, C. A. & Turvey, M. T. 2006 The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377.

Ganis, G., Kutas, M. & Sereno, M. I. 1996 The search for 'common sense': an electrophysiological study of the comprehension of words and pictures in reading. *J. Cogn. Neurosci.* **2**, 89–106. (doi:10.1162/jocn.1996.8.2.89)

Glanemann, R., Reichmuth, K., Fiori, A., am Zehnhoff-Dinnesen, A. & Dobel, C. 2008 Computerbasiertes Verblernen bei Kindern mit Cochlea Implantat. In *Aktuelle phoniatrisch-pädaudiologische Aspekte*, vol. 16 (eds M. Gross & A. am Zehnhoff-Dinnesen), pp. 141–144. Mönchengladbach, Germany: Rheinware Verlag.

Goodsit, J. V., Morgan, J. L. & Kuhl, P. K. 1993 Perceptual strategies in prelingual speech segmentation. *J. Child Lang.* **20**, 229–252. (doi:10.1017/S0305000900008266)

Guillem, F., N'Kaoua, B., Rougier, A. & Claverie, B. 1996 Differential involvement of the human temporal lobe structures in short- and longterm memory processes assessed by intracranial ERPs. *Psychophysiology* **33**, 720–730. (doi:10.1111/j.1469-8986.1996.tb02368.x)

Hagoort, P., Brown, C. M. & Swaab, T. Y. 1996 Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia and patients with right hemisphere lesions without aphasia. *Brain* **119**, 627–649. (doi:10.1093/brain/119.2.627)

Hämäläinen, M. S. & Ilmoniemi, R. J. 1994 Interpreting magnetic fields of the brain—minimum-norm estimates. *Med. Biol. Eng. Comput.* **1**, 35–42. (doi:10.1007/BF02512476)

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S. & Lyytinen, H. 2002 Cortical activation during spoken word segmentation in nonreading impaired and dyslexic adults. *J. Neurosci.* **22**, 2936–2944.

Hillenbrand, J. 1984 Speech perception by infants: categor-ization based on nasal consonant place of articulation. *J. Acoust. Soc. Am.* **5**, 1613–1622. (doi:10.1121/1.390871)

Holcomb, P. J. & McPherson, W. B. 1994 Event-related brain potentials reflect semantic priming in an object decision task. *Brain Cogn.* **2**, 259–276. (doi:10.1006/brcg.1994.1014)

Holcomb, P. J. & Neville, H. J. 1990 Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Proc.* **5**, 281–312. (doi:10.1080/01690969008407065)

Kissler, J., Herbert, C., Peyk, P. & Junghofer, M. 2007 Buzz-words: early cortical responses to emotional words during reading. *Psych. Sci.* **18**, 475–480. (doi:10.1111/j.1467-9280.2007.01924.x)

Knecht, S., Breitenstein, C., Bushuven, S., Wailke, S., Kamping, S., Flöel, A., Zwitserlood, P. & Ringelstein, E. B. 2004 Levodopa: faster and better word learning in

normal humans. *Ann. Neurol.* **56**, 20–26. (doi:10.1002/ana.20125)

Kuhl, P. K. 1979 Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *J. Acoust. Soc. Am.* **66**, 1668–1679. (doi:10.1121/1.383639)

Kuhl, P. K. 1991 Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* **2**, 93–107.

Kuhl, P. K. 2004 Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **11**, 831–841. (doi:10.1038/nrn1533)

Kuhl, P. K., Williams, K. A., Lacerda, F. & Stevens, K. N. 1992 Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **5044**, 606–608. (doi:10.1126/science.1736364)

Kuhl, P. K., Tsao, F. M., Liu, H. M., Uhang, Y. & De Boer, B. 2001 Language/culture/mind/brain. Progress at the margins between disciplines. *Ann. NY Acad. Sci.* **935**, 136–174.

Kutas, M. & Hillyard, S. A. 1980 Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* **4427**, 203–205. (doi:10.1126/science.7350657)

Kutas, M., Neville, H. J. & Holcomb, P. J. 1987 A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing. *Electroencephalogr. Clin. Neurophysiol.* **39**(Suppl.), 325–330.

Kwon, H., Kuriki, S., Kim, J. M., Lee, Y. H., Kim, K. & Nam, K. 2005 MEG study on neural activities associated with syntactic and semantic violations in spoken Korean sentences. *Neurosci. Res.* **51**, 349–357. (doi:10.1016/j.neures.2004.12.017)

Lasky, R. E., Syrdal-Lasky, A. & Klein, R. E. 1975 VOT discrimination by four to six and a half month old infants from Spanish environments. *J. Exp. Child Psychol.* **2**, 215–225. (doi:10.1016/0022-0965(75)90099-5)

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. 1967 Perception of the speech code. *Psychol. Rev.* **74**, 431–461. (doi:10.1037/h0020279)

Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A. & Friederici, A. 2006 Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Res.* **22**, 163–172. (doi:10.1016/j.brainres.2006.04.037)

McLaughlin, J., Osterhout, L. & Kim, A. 2004 Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* **7**, 703–704. (doi:10.1038/nn1264)

Menning, H., Imaizumi, S., Zwitserlood, P. & Pantev, C. 2002 Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learn. Mem.* **9**, 253–267. (doi:10.1101/lm.49402)

Mestres-Missé, A., Rodriguez-Fornells, A. & Münte, T. 2007 Watching the brain during meaning acquisition. *Cereb. Cortex* **17**, 1858–1866. (doi:10.1093/cercor/bhl094)

Miller, J. L. & Liberman, A. M. 1979 Some effects of later-occurring information on the perception of stop consonant and semivowel. *Percept. Psychophys.* **6**, 457–465.

Nobre, A. C. & McCarthy, G. 1995 Language-related field potentials in the anterior-medial temporal lobe. II. Effects of word type and semantic priming. *J. Neurosci.* **15**, 1090–1098.

Nobre, A. C., Allison, T. & McCarthy, G. 1994 Word recognition in the human inferior temporal lobe. *Nature* **372**, 260–263. (doi:10.1038/372260a0)

Plante, E., Petten, C. V. & Senkfor, A. J. 2000 Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia* **13**, 1669–1684. (doi:10.1016/S0028-3932(00)00083-X)

Pulvermüller, F. & Shtyrov, Y. 2006 Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* **1**, 49–71. (doi:10.1016/j.pneurobio.2006.04.004)

Rivera-Gaxiola, M., Silva-Pereyra, J. & Kuhl, P. K. 2005 Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Dev. Sci.* **2**, 162–172. (doi:10.1111/j.1467-7687.2005.00403.x)

Rugg, M. D. 1985 The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology* **6**, 642–647. (doi:10.1111/j.1469-8986.1985.tb01661.x)

Saffran, J. R. 2002 Constraints on statistical language learning. *J. Mem. Lang.* **1**, 172–196. (doi:10.1006/jmla.2001.2839)

Saffran, J. R., Aslin, R. N. & Newport, E. L. 1996 Statistical learning by 8-month old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)

Schomacher, M., Baumgaertner, A., Winter, B., Lohmann, H., Dobel, C., Wedler, K., Abel, S., Knecht, S. & Breitenstein, C. 2006 Erste Ergebnisse zur Effektivität eines intensiven und hochfrequent repetitiven Benenn- und Konversationstrainings bei Aphasie. *Forum Logopädie* **20**, 22–28.

Schupp, H. T., Stockburger, J., Codispoti, M., Junghofer, M., Weike, A. I. & Hamm, A. O. 2007 Selective visual attention to emotion. *J. Neurosci.* **5**, 1082–1089. (doi:10.1523/JNEUROSCI.3223-06.2007)

Sebastian-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R. & Díaz, B. 2006 First- and second-language phonological representations in the mental lexicon. *J. Cogn. Neurosci.* **18**, 1277–1291. (doi:10.1162/jocn.2006.18.8.1277)

Simos, P. G., Basile, L. F. H. & Papanicolaou, A. C. 1997 Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Res.* **762**, 29–39. (doi:10.1016/S0006-8993(97)00349-1)

Sjerps, M. J. & McQueen, J. M. In press. The bounds on flexibility in speech perception. *J. Exp. Psychol. Hum. Percept. Perform.*

Smith, M. E., Stapleton, J. M. & Halgren, E. 1986 Human medial temporal lobe potentials evoked in memory and language tasks. *Clin. Neurophysiol.* **63**, 145–159. (doi:10.1016/0013-4694(86)90008-8)

Strevens, P. 1960 Spectra of fricative noise in human speech. *Lang. Speech* **3**, 32–49.

Swaab, T., Brown, C. & Hagoort, P. 1997 Spoken sentence comprehension in aphasia: event-related potential evidence for a lexical integration deficit. *J. Cogn. Neurosci.* **9**, 39–66. (doi:10.1162/jocn.1997.9.1.39)

Van Petten, C. 1993 A comparison of lexical and sentence-level context effects in event-related potentials. *Lang. Cogn. Proc.* **4**, 485–531. (doi:10.1080/01690969308407586)

Van Petten, C. & Kutas, M. 1990 Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cognit.* **4**, 380–393.

Van Petten, C. & Kutas, M. 1991 Influences of semantic and syntactic context in open- and closed-class words. *Mem. Cognit.* **1**, 95–112.

Van Petten, C. & Luka, B. J. 2006 Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang.* **3**, 279–293. (doi:10.1016/j.bandl.2005.11.003)

Van Petten, C. & Rheinfelder, H. 1995 Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. *Neuropsychologia* **4**, 485–508. (doi:10.1016/0028-3932(94)00133-A)

Werker, J. F. & Lalonde, C. E. 1988 Cross-language speech perception: initial capabilities and developmental change. *Dev. Psychol.* **5**, 672–683. (doi:10.1037/0012-1649.24.5.672)

Werker, J. F. & Tees, R. C. 1984 Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev. Int. Interdiscip. J.* **1**, 49–63. (doi:10.1016/S0163-6383(84)80022-3)

Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A. & Carr, T. H. 1997 Word, pseudoword, and nonword processing: a multitask comparison using event-related brain potentials. *J. Cogn. Neurosci.* **9**, 758–775. (doi:10.1162/jocn.1997.9.6.758)