

Journal Pre-proof

In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects

Samuel H. Cospers, Claudia Männel, Jutta L. Mueller



PII: S1878-9293(20)30069-4
DOI: <https://doi.org/10.1016/j.dcn.2020.100821>
Reference: DCN 100821

To appear in: *Developmental Cognitive Neuroscience*

Received Date: 18 June 2019
Revised Date: 13 May 2020
Accepted Date: 29 June 2020

Please cite this article as: Cospers SH, Männel C, Mueller JL, In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects, *Developmental Cognitive Neuroscience* (2020), doi: <https://doi.org/10.1016/j.dcn.2020.100821>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects

Authors: Samuel H. Cospers^{1*}, Claudia Männel^{2,3}, & Jutta L. Mueller¹

Affiliations:

¹Institute of Cognitive Science, University of Osnabrück

²Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig

³Department of Audiology and Phoniatics, Charité–Universitätsmedizin Berlin

*Corresponding author at: Institute of Cognitive Science, University of Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany; *email address:* samuel.cospers@uni-osnabrueck.de (S. Cospers);

Emails: samuel.cospers@uni-osnabrueck.de (S. Cospers); maennel@cbs.mpg.de (C. Männel); jutta.mueller@uni-osnabrueck.de (J. Mueller)

Highlights:

- ERPs during training showed word-form a familiarity effect over time
- ERPs at training showed a pairing consistency effect over time
- ERPs at test showed lateralized N400-like responses to violated sound-word pairs
- Infants can learn novel words for sounds in a similar way as for visual objects

Abstract:

Despite the prominence of non-visual semantic features for some words (e.g., siren or thunder), little is known about when and how the meanings of those words that refer to auditory objects can be acquired in early infancy. With associative learning being an important mechanism of word learning, we ask the question whether associations between sounds and words lead to similar learning effects as associations between visual objects and words. In an event-related potential (ERP) study, 10- to 12-month-old infants were presented with pairs of environmental sounds and pseudowords in either a consistent (where sound-word mapping can occur) or inconsistent manner. Subsequently, the infants were presented with sound-pseudoword combinations either matching or violating the consistent pairs from the training phase. In

the training phase, we observed word-form familiarity effects and pairing consistency effects for ERPs time-locked to the onset of the word. The test phase revealed N400-like effects for violated pairs as compared to matching pairs. These results indicate that associative word learning is also possible for auditory objects before infants' first birthday. The specific temporal occurrence of the N400-like effect and topological distribution of the ERPs suggests that the object's modality has an impact on how novel words are processed.

Keywords: word learning, associative learning, Event-related Potential, EEG, language acquisition, auditory modality

1 Introduction

Our daily lives are full of sensory input from different modalities and language provides us with the means to describe those rich experiences, with words for visual, auditory, olfactory and haptic experiences such as *flower* (visual), *thunder* (auditory), *stench* (olfactory), and *tickle* (haptic). Despite this rich and multimodal environment, research on how we acquire words for entities and events has mostly focused on labeling objects in the visual modality (e.g., Borgström, Torkildsen, & Lindgren, 2015; Friedrich & Friederici, 2008; Horst & Samuelson, 2008; Junge, Cutler, & Hagoort, 2012; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Smith & Yu, 2008; Taxitari, Twomey, Westermann, & Mani, 2019; Torkildsen et al., 2008; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007; Werker, Cohen, Lloyd, Casasola, & Stager, 1998). This imbalance may originate from the larger number of words for visible things in many languages. A recent study has found that adult speakers of English relate a higher number of nouns to the visual modality than to other modalities, but that the auditory modality has the second highest number of nouns allocated to it (Winter, Perlman, & Majid, 2018). Furthermore, it has also been recently discovered that the hierarchy of these perceptual modalities differs between languages indicating the necessity to investigate word learning mechanisms in modalities other than the visual modality (Majid et al., 2018). In order to explore whether and how words for non-visual objects can be acquired in infancy, we investigated 10- to 12-month-old infants' ability to map novel labels onto auditory objects.

To acquire the meaning of a word, a connection between an object and its respective label must be established. This happens over time and typically through repeated exposure to or experience with an object and its label. Within the scope of the current study, we define word learning as an associative process, during which the relationship between co-occurring objects and labels becomes stronger over time and with experience (McMurray, Horst, & Samuelson, 2012; Sloutsky, Yim, Yao, & Dennis, 2017). Even if associative learning is a single mechanism, it has been shown to explain different phenomena in word learning, including for example the progression from slow to fast learning of word meanings (cf.

Regier, 2005). Note, though, that alternative mechanisms have been suggested as well, for example inference-based learning, in which learners initially consider a plethora of possible concepts that can map onto a word, which eventually converge to a single hypothesis with growing experience (for a short review, see Sloutsky, Yim, Yao, & Dennis, 2017). Within the scope of the current study, however, we propose that associative learning is sufficient for building a first link between an object and a word, particularly during online learning and short term retention without generalization (cf. McMurray et al., 2012).

In an environment rich with complex and concurrent sensory input, associating a word with a specific object is a non-trivial task that infants start to approach from early on. Both behavioral (e.g., head turning and preferential looking) and electrophysiological studies have shown that infants are sensitive to word-form familiarization and capable of word learning in laboratory settings by means of frequent word-object co-occurrences (for a review, see Johnson, 2016). In behavioral studies on word knowledge in the native language, evidence of a link between visual objects and their labels (that were both highly frequent in the infants' environment, such as *spoon* or *banana*) has already been observed at six months of age (e.g., Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012). Laboratory learning experiments using behavioral measures also show first evidence of infant word learning around six months of age (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Johnson, Seidl, & Tyler, 2014; Shukla, White, & Aslin, 2011). In electrophysiological experiments, evidence for differences in processing the correct and incorrect use of real words for visual objects is not found before 14 months of age (Friedrich & Friederici, 2005; Mills et al., 2004); however, electrophysiological evidence for online association-building between visual objects and pseudowords, has been reported starting from the age of three months (Friedrich & Friederici, 2017). Given these differences in developmental timelines between studies, it seems that behavioral and electrophysiological methods might capture partially different aspects of word learning. While behavioral tests are sensitive in the assessment of already existing word knowledge, electrophysiological methods can additionally capture online learning processes in lab-based learning designs. Despite potential differences in sensitivity to capture specific processes across development, all studies do show that early word knowledge, or word-object association, is facilitated by highly frequent occurrences of the respective visual objects in everyday life.

As reviewed above, previous research on infant word learning has so far only focused on the visual modality, testing associative auditory-visual word learning, while uni-modal associative word learning in the auditory modality has not been captured. This dominance of the visual modality might be partly explained by research on early word learning focusing on objects' perceptual saliency in the infant's focus of attention (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Samuelson & Smith, 1998; Smith, Jones, & Landau, 1996). For example, Smith and colleagues provided evidence that the visual input from an egocentric view plays a major role for word learning (Smith,

Jayaraman, Clerkin, & Yu, 2018; Yu & Smith, 2012). Yet, these studies leave aside the possibility that perceptual input from other modalities could serve as a salient input for word learning as well.

Auditory input might be a salient source of sensory information in infant word learning, as auditory sensitivity can be seen already in utero (Kisilevsky et al., 2003) and in neonates (DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Demany, 1982; Demany, McKenzie, & Vurpillot, 1977; Panneton & DeCasper, 1986). Visual processing, on the other hand, requires several months after birth in order to reach the level of maturity of auditory processing (Banks & Salapatek, 1983). Further, infants and children have a general preference toward the auditory modality, also known as auditory dominance, that was found in 6- and 10-month-olds (Lewkowicz, 1988a, 1988b). In the light of the early onset of auditory perception it is no surprise that infants are capable of applying statistical learning to auditory sequences early in life (Richards & Goldfarb, 1986; Saffran, 2002; Saffran, Aslin, & Newport, 1996; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Thus, research in young infants shows that auditory perception matures early, might be preferred over visual perception and can be exploited by associative (statistical) learning processes. Interestingly, further investigations have shown that the auditory preference is still present in later childhood at four years of age, yet ultimately disappears in adulthood, where adults exhibit a visual preference (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). In light of the advantageous status of the auditory modality in early infancy it is notable that the majority of electrophysiological research on early word learning has been conducted in cross-modal experiments (e.g., Borgström, Torkildsen, & Lindgren, 2016, 2015a, 2015b, Friedrich & Friederici, 2008, 2011, 2017; Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Torkildsen et al., 2008, 2009). Thus, by exploring paradigms presenting objects in other than the visual modality, for example, audition, we can begin to understand how different types of perceptual information affect the building of object-word associations early in life, such as the meanings for auditory words like *lullaby*, *siren*, and *thunder*.

In order to test infants' associative word learning capabilities in the current study, we used the event-related potential (ERP) technique to determine changes in neural processing over time. Although there are many behavioral procedures that also test infants' associative learning (e.g., eye-tracking; Yu, Zhong, & Fricker, 2012 and preferential looking; Bergelson & Aslin, 2017), we chose the ERP technique, as it can measure association and semantic learning processes independent of behavioral responses. The particular ERP component associated with the semantic expectation of a given word in a particular context is the N400, which is characterized by a more negative ERP response to a contextually violated condition than to a contextually expected control condition (Kutas & Hillyard, 1980). Since its discovery, the N400 component has been found in many experiments on semantic priming, including language/word-learning in adults (e.g. Bentin, McCarthy, & Wood, 1985; Boddy, 1986; Kutas & Federmeier, 2011; Kutas, Lindamood, & Hillyard, 1984) and children (e.g., Borgström, Torkildsen, & Lindgren, 2016, 2015a, 2015b, Friedrich & Friederici, 2008, 2011, 2017; Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Torkildsen et al., 2008, 2009). Note, that the N400

component does not only occur in the context of spoken and written words, but also for semantic processing difficulties in non-linguistic materials (Cummings et al., 2006; Koelsch et al., 2004).

Although the N400 has been reported to reflect semantic priming, the functional interpretation of the component is still under discussion. The two main interpretations are first, spreading activation (Posner & Snyder, 1975), an automatic process where activation is forwarded from a prime (e.g., object) to associated items (e.g., the target word); and second, semantic integration, a process of relating the prime to the target in order to form a combined meaning (for a review, see Kutas & Federmeier, 2011). Although some findings report on the N400 only reflecting semantic integration (Bentin, Kutas, & Hillyard, 1993, 1995), other studies have suggested spreading activation as a likely mechanism (Holcomb, 1988; Kiefer, 2002; Lau, Phillips, & Poeppel, 2008). Hence, the presence of the N400 priming effect can be interpreted as evidence for both automatic and more controlled lexical-semantic processes. With respect to the spreading activation account, it is interesting to note that N400 effects are even reported for non-semantic associative learning (Ortu, Allan, & Donaldson, 2013; Rhodes & Donaldson, 2007; Tabullo, Yorio, Zanutto, & Wainseboim, 2015).

Infants as young as 6- to 9-months-old have been shown, via ERPs, to have a sensitivity to visual object-word associations (e.g., Friedrich & Friederici, 2008, 2011; Junge, Cutler, & Hagoort, 2012; Parise & Csibra, 2012). Several ERP components have been found which indicate, potentially at different representation levels, word-form familiarity effects or violated expectation effects in word learning studies: the N200-500 (Friedrich & Friederici, 2008; Kooijman, Hagoort, & Cutler, 2005; Mills, Coffey-Corina, & Neville, 1993; Thierry, Vihman, & Roberts, 2003; Torkildsen et al., 2009) and the N400-like component (Borgström et al., 2015b; Friedrich & Friederici, 2008; Junge et al., 2012; Torkildsen et al., 2008, 2007), respectively. In infants, word-form familiarity effects occur for repeated stimulus presentation or already learned versus novel words and have been reported for both real words and pseudowords (e.g., Borgström et al., 2015b; Friedrich & Friederici, 2011, 2017; Junge et al., 2012; Kooijman et al., 2005; Männel & Friederici, 2013; Obrig et al., 2017; Torkildsen et al., 2009). The effect is typically found around 200-500 ms (for a review, see Teixidó, François, Bosch, & Männel, 2019), but also from 400-1200 ms (Junge, Cutler, & Hagoort, 2014; Torkildsen et al., 2009), after the onset of the word across fronto-central regions as measured by an increased negative amplitude for repetition over time or familiarity. The N400 has also been used as a measure for associative word learning in infant studies and reflects a violation of contextual expectation, with more negative amplitude waveforms for violated than matching conditions. As its timing and topography do not always correspond to the properties of the adult N400, we term it N400-like component in the context of infant studies (cf. Junge et al., 2012).

The present study aims to explore whether novel words can be associated with objects in the auditory modality in young infants. We used a uni-modal auditory associative-learning paradigm (cf. Friedrich & Friederici, 2008, 2011, 2017). Based on the findings of Saffran (2002), we presented auditory objects

and pseudowords sequentially, as infants' statistical learning benefits from sequential stimulus presentation in the auditory modality. Particularly, we aimed to evaluate electrophysiological effects in 10- to 12-month-olds during a training phase, where pseudowords were consistently or inconsistently paired with auditory objects, and a subsequent testing phase, where the previously encountered pairings were either matched or violated. This age group was selected as 10- to 12-month-olds are expected to still have a strong auditory preference (cf. Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003) and thus experience environmental sounds as salient input. Specifically, we expect to see a *word-form familiarity effect* in the ERP, that is, an enhanced negativity for word-form repetitions irrespective of the consistency of the sound-word pairings in the training phase. Additionally, we expect ERP results to show a priming effect for the pseudowords in the training phase when comparing the first four presentations of the sound-word pairing to the final four presentations. Here, we predict a *pairing consistency effect*, such that the ERP amplitude to pseudowords of consistently paired stimuli will reduce over time, while the response to the inconsistent pairings remains more constant (cf. Friedrich & Friederici, 2011). For the testing phase, we expect an *N400-like effect* for pseudowords in violated as compared to matching sound-word pairings, reflecting the violation of a lexical-semantic expectation. Furthermore, we will descriptively compare the ERP effects observed in our study with ERP effects reported in previous visual-auditory association paradigms (e.g., Borgström et al., 2015a, 2015b; Friedrich & Friederici, 2008, 2011; Parise & Csibra, 2012).

2 Materials and methods

2.1 Participants

A group of 55 infants (32 boys, 23 girls) was tested between 10 and 12 months of age. The datasets of 32 infants were included in the final analysis (19 boys and 13 girls; mean age = 10 months, 24 days; $SD = 17.4$ days). Datasets of 18 infants were excluded due to fussiness and thus, having too many motion-related artifacts in the EEG. Four further datasets were excluded due to failure to complete the experiment, and one was excluded for non-compliance in wearing the EEG cap. All children were typically developing, carried to at least 37 weeks of gestation, were learning German as a native language, had normal hearing, and no reported neurological conditions or learning/language impairments. Families were contacted via the database of the Max Planck Institute for Human Cognitive and Brain Sciences (MPI-CBS) in Leipzig, Germany or the database of the Institute of Cognitive Science at the University of Osnabrück in Osnabrück, Germany. Written informed consent was given by the parents before their child's participation. The families were given a monetary travel reimbursement and the infants were given a gift for their participation, totaling to about 17 euro. The study was approved by the ethics committees of the Medical Faculty of the University of Leipzig and the University of Osnabrück.

2.2 Stimuli

The material presented during the experiment consisted of auditory material (see supplementary files for a complete list of all stimuli), combined in pairs of auditory objects and pseudoword labels. The auditory object stimuli consisted of 16 environmental sounds (e.g., boiling water, crackling fire, various animal sounds) and were taken from the NESSTI database for environmental sounds (Hocking, Dzafic, Kazovsky, & Copland, 2013). The environmental sounds selected for the experiment were taken from larger audiofiles and cut to 950 ms. The labels consisted of 16 disyllabic pseudowords that are phonetically legal in German and derived from existing German nouns. The pseudowords were recorded with natural intonation by a female native speaker of German. The environmental sounds were 950 ms in length and pseudowords varied between 650 to 750 ms, but silence was added to the end of the word files to make each .WAV file 750 ms in length. All stimuli were saved as monaural sound files (duplicated over both channels), and digitized at 44100 Hz.

2.3 Procedure

The experiment was divided into two phases: an initial training phase, which was immediately followed by a subsequent testing phase. In the training phase, infants were presented sound-word combinations of the auditory stimuli. The combinations were divided equally into consistent pairings and inconsistent pairings, counter-balanced across children. The consistent pairings were comprised of eight sounds each consistently paired with one of eight pseudowords; in this condition, associative learning was possible. Each pair was presented eight times for a total of 64 trials. For the inconsistent pairings, the remaining eight sounds and pseudowords were distributed in a rotated manner so that each sound was presented with each word exactly once for a total of 64 trials; in this condition, associative learning was not possible. Figure 1A gives a visual representation of the sound-pseudoword distribution and presentation in the training phase. The testing phase consisted of two conditions: matching pairs and violated pairs. The matching condition corresponded to the consistent pairings of the training phase. Violations consisted of sounds and pseudowords from the consistent pairings, but rearranged in a manner as to not match the sound-pseudoword pairings of the training phase (see Figure 1B). Each condition in the testing phase consisted of eight items, presented twice, totaling in 16 trials per condition. Eight separate stimuli lists assured randomized stimulus presentations across participants by switching the consistent and inconsistent sound-pseudoword pairs, reshuffling the distribution of sound-pseudoword pairs, and backward presentation of all lists to account for presentation order. Each list was presented to four participants in the final analysis group.

[Figure 1 here]

The structure of all trials was identical. First, an environmental sound was presented for 950 ms followed by a pause of 600 ms. Subsequently, a pseudoword was presented. The time between trials was 1600 ms in order to acoustically indicate individual pairings (see Figure 1C). Total experiment time was 12 minutes and 20 seconds. During this time, the infants sat on their parent's lap facing a computer monitor with speakers on either side of the monitor. In order to raise infants' compliance, a silent cartoon was played on the screen on a smaller scale in order to keep eye movements minimal (for similar procedures, see Männel, Schaadt, Illner, van der Meer, & Friederici, 2017; Männel, Schipke, & Friederici, 2013). Additionally, a second experimenter sat near the child to be able to further entertain the child with either silent toys, bubbles, or a picture book if necessary.

2.4 EEG Processing

EEG data were collected in two different locations: at the MPI-CBS, Leipzig and the University of Osnabrück. Both locations used the EEG amplifier REFA (Twente Medical Systems International, Oldenzaal, The Netherlands) and all participants were tested with the same elastic caps with 27 implemented Ag/AgCl electrodes (EASYCAP GmbH, Herrsching, Germany) at the standard 10-20 electrode positions Fp1/2, Fz, F3/4, F7/8, F9/10, FC5/6, Cz, C3/4, T7/8, CP5/6, TP9/10, Pz, P3/4, P7/8, O1/2 (ground electrode located at AFz). Electrode impedances were kept below 30 k Ω . For the infants tested at the MPI-CBS ($N = 8$), EEG data were continuously recorded using the QRefa Acquisition Software, Version 1.0 beta (MPI-CBS, Leipzig, Germany) with a sampling rate of 500 Hz, an online reference at Cz, and a monopolar electrooculogram (EOG) electrode under the right eye. For the infants tested at the University of Osnabrück ($N = 24$), the EEG data were recorded with the TMSi Polybench Software (Twente Medical Systems International, Oldenzaal, The Netherlands) at a sampling rate of 512 Hz, an average online reference, and a monopolar EOG electrode under the left eye.

All data analyses and statistics were conducted in MATLAB (The Mathworks Inc., Natick, MA, USA) and EEG data were processed using EEGLAB (Delorme & Makeig, 2004). All EEG epochs time-locked to the onset of the pseudowords were created with a length of 1200 ms and a baseline of 200 ms pre-stimulus. The EEG data were re-referenced to the linked mastoids (electrodes at sites TP9/10). After initial artifact rejection, using both automatic marking of min-max voltage changes of 100 μ V and manual inspection, the epochized datasets of participants were individually, semi-automatically corrected for eye movements using an EEGLAB independent component analysis (ICA). For the ICA, a high-pass filter of 1 Hz (-3dB, cutoff frequency of 1.38 Hz) was applied to the continuous EEG dataset of each participant. The ICA weights from the 1 Hz-filtered dataset were then applied to a copy of the dataset with a high-pass filter of 0.3 Hz (-3dB, cutoff frequency of 0.36 Hz). After applying the ICA weights, artifact rejection for the 0.3 Hz filtered dataset was conducted using same semi-automatic process described above. All subsequent data analyses and statistics were conducted using the 0.3 Hz filtered dataset. Before creating the grand-average across all participants, a low-pass filter of 25 Hz (-3dB, cutoff frequency of 27.35 Hz) was applied. The final dataset for consistently paired pseudowords

in the first half of the training phase consisted of an average of 17.4 out of 32 trials ($SD = 5.8$ trials) and 16.5 out of 32 trials ($SD = 5.5$ trials) for the inconsistently paired pseudowords. For the second half of the training phase, the final dataset for the consistently paired pseudowords consisted of an average of 15.2 out of 32 trials ($SD = 5.4$ trials) and 14.5 out of 32 trials ($SD = 5.4$ trials) for the inconsistently paired pseudowords. For the testing phase, the final dataset for the matched pseudowords included an average of 6.8 out of 16 trials ($SD = 2.9$ trials) and 6 out of 16 trials ($SD = 2.6$ trials) for the violated pseudowords.

2.5 Statistical Analysis

For the statistical analysis, 15 electrodes were chosen and combined in Regions of Interest (ROIs) for laterality (left: F7, T7, P7; left medial: F3, C3, P3; medial: Fz, Cz, Pz; right medial: F4, C4, P4; right: F8, T8, P8) and region (anterior: F7, F3, Fz, F4, F8; central: T7, C3, Cz, C4, T8; posterior: P7, P3, Pz, P4, P8).

In order to test for the familiarization of the pseudowords (*word-form familiarity effect*) and object-word association over time (*pairing consistency effect*) during the training phase, a $5 \times 3 \times 2 \times 2$ repeated-measures ANOVA was calculated using the mean ERP amplitudes with the within-subject factors of laterality (left, mid-left, midline, mid-right, right), region (anterior, central, posterior), repetition (1st to 4th presentation vs. 5th to 8th presentation)¹, and consistency (consistent vs. inconsistent pairing).

In the test phase, analyses of ERP responses evoked by pseudowords in violated compared to matching sound-word pairs (*N400-like effect*) should reveal whether infants had successfully associated sounds and pseudowords during the training phase. This was captured in a $5 \times 3 \times 2$ repeated-measures ANOVA using the mean ERP amplitudes with the within-subject factors of laterality (left, mid-left, midline, mid-right, right), region (anterior, central, posterior), and condition (matching pairs vs. violated pairs).

The time windows (TWs) of interest in both the training and testing phases were set at 100 ms intervals starting from the onset of the stimulus until the end of the epoch (1200 ms). As multiple individual repeated-measures ANOVAs were calculated for each phase, the significance levels of all p -values were corrected using the Bonferroni-correction method. A post-hoc step-down analysis was performed for all interactions involving either the factor repetition, consistency or condition where the p -value was $p \leq 0.1$. For all ANOVAs, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied and corrected p -values and uncorrected degrees of freedom are presented.

¹ The repeated-measures ANOVA compares the first half of the training phase with the second half of the training phase and thus provides a proof of principle test for the effect of repetition. With a more fine-grained repetition factor, the number of trials included per factor level would be too few, resulting in very low signal-to-noise ratio.

3 Results

3.1 Training Phase

The ERP data show evidence for a modulation of the ERP waveforms across both halves of the training phase: First, the pseudowords presented in the second half of the training phase elicited a more negative-going waveform over anterior and central sites compared to those presented in the first half (i.e., *word-form familiarity effect* that is independent of the pairing condition) (see Figure 2). Second, consistently combined pseudowords evoked a more negative-going waveform than inconsistently combined pseudowords in the second half of the training phase (i.e., *pairing consistency effect* that only arises later during the experiment) (see Figure 3). In the following, only significant or marginally significant effects will be reported.

The 5x3x2x2 repeated-measures ANOVAs revealed a significant main effect for repetition, i.e. a *word-form familiarity effect*, in the 600-700 ms TW ($F(1.00,31.00)=14.029$, $p = 0.001$). Significant interactions were found between repetition and region spanning from 400 to 700 ms after the onset of the pseudoword, giving evidence to effects of *word-form familiarity* with an anterior-central distribution. See Table 1 for the significant repetition by region interaction and step-down analyses.

[Table 1 here]

[Figure 2 here]

Further, a *pairing consistency effect*, by means of a marginally significant interaction between laterality, region, repetition, and consistency, was found in the 500-600 ms TW after the onset of the pseudoword ($F(5.70,176.79)=3.058$, $p = 0.008$). Step-down analyses revealed for the first half of the training phase, a significant consistency effect right temporally (T8: $T(31) = -2.2981$, $p = 0.0285$), and for the second half, significant consistency effects left central-laterally (T7: $T(31) = -2.1443$, $p = 0.04$; C3: $T(31) = -3.1293$, $p = 0.0038$; Cz: $T(31) = -2.4715$, $p = 0.0192$; Pz: $T(31) = -2.1349$, $p = 0.0408$).

[Figure 3 here]

3.2 Testing Phase

The ERP data in the testing phase revealed a lateralized increased negativity in the violation versus matching conditions with an *N400-like effect* (see Figure 4). For the TW 300-400 ms after the onset of the word, a condition by laterality interaction was observed $F(3.12,96.69) = 5.710$, $p = 0.001$. In step-down analyses, condition effects were found to be left-lateralized (left: $T(31) = 2.079$, $p = 0.046$; medial left: $T(31) = 2.361$, $p = 0.025$; and on the midline: $T(31) = 2.982$, $p = 0.0055$).

[Figure 4 here]

4 Discussion

The aim of the present study was to test whether 10- to 12-month-old infants are capable of mapping novel labels onto auditory objects in the absence of object-related concurrent visual input. We found ERP evidence for both familiarization of word-forms across the training phase as well as for the learning of sound-label associations in the training phase. Importantly, in the subsequent testing phase, infants showed recognition of sound-label violations of the previously established associations.

More specifically, in the training phase, we observed two ERP effects: First, the *word-form familiarity effect* that occurred at fronto-central regions spanning from 400-700 ms after the onset of the pseudoword. This ERP effect displayed as more negative-going responses in the second half of the training phase than in the first half. Second, we observed a *pairing consistency effect* at the right-temporal region in the 500-600 ms TW with more negative-going responses to consistently paired pseudowords than to inconsistently paired pseudowords. Here, we hypothesized that the inconsistently paired pseudowords would be more negative than the consistently paired pseudowords; however, we found the opposite. Interestingly, this pairing consistency effect already occurred during the first half of the experiment, yet had a broader, left centro-temporal distribution in the second half. This suggests that infants seem to be capable of associating sound-pseudoword pairs even within only four presentations and that this association then might become stronger over time.

In the testing phase, we observed an effect of violated expectation, that is the ERPs were more negative-going for pseudowords in violated than matched sound-pseudoword pairs. This effect occurred left-lateralized in the 300-400 ms TW. We interpret this effect as an *N400-like effect*; the more negative responses to violated pairs suggest an association has been formed for the consistent sound-word pairs in the training phase, which then was unexpectedly not met. Thus, infants are not only able to form associations for sound-pseudoword pairs, but are also able to recognize violations to established pairs in a subsequent test phase.

As hypothesized for the training phase, we found an effect of pseudoword repetition that occurs irrespective of the consistency of the paired sound and thus indicates that the pseudowords have been familiarized over time. This *word-form familiarity effect* evolved in the second half of the training phase at 400-700 ms after pseudoword onset. We argue that this effect is a replication of the familiarity effect found in cross-modal paradigms, although this previous effect has been typically reported for the 200-500 ms TW (e.g., Friedrich & Friederici, 2011; Kooijman et al., 2005; Torkildsen et al., 2009). One plausible factor that could contribute to the later occurrence of the *word-form familiarity effect* in our study is the extra processing due to the presentation of both speech and non-speech acoustic stimuli.

Junge, Cutler, and Hagoort (2014) reported similar delayed word-form familiarity effects for 10-month-old infants, at 350-500 ms and 600-900 ms in a word-segmentation experiment, suggesting that the processing of novel words in sentence contexts is more difficult than the mapping of known words to existing memory traces. Similarly, Obrig and colleagues (2017) conducted a cross-modal word-learning study with 6-month-olds which revealed a larger negativity for pseudowords after training in the 450-850 ms TW. Thus, the timing of the word-form familiarity effect seems to depend on context and stimulus type. We argue that it could be delayed in the current study due to processing costs attributed to the discrimination of linguistic (label) versus non-linguistic (sound) acoustic stimuli.

In addition, we also observed the postulated *pairing consistency effect* in the second half of the training phase, similarly to infant studies, in which visual objects were paired with spoken words. In contrast to our study, however, these cross-modal studies revealed N400-like effects, with amplitude reductions for consistent versus inconsistent object-word pairs (e.g., Borgström et al., 2015b; Friedrich & Friederici, 2008), indicating that semantic or at least cross-modal associations are formed after repeated presentations of identical pairs. Interestingly, a *pairing consistency effect* with a reversed polarity, comparable to our study, was only shown in the youngest of the tested age groups, for associative cross-modal learning at 3 months (Friedrich & Friederici, 2017). In their study, the authors found an increased negativity in the second half of the training phase for consistently versus inconsistently paired pseudowords (with visual objects) at the left centro-parietal site from 500-1000 ms after the onset of the pseudoword. While such negativities are different from the typically reported semantic priming N400-like effects, we take this *pairing consistency effect* to show that infants gradually associate consistently paired pseudowords with the related sounds across experimental time.

The interpretation of the *pairing consistency effect* as an indication of successful associative learning is supported by studies showing enhanced N400 effects in response to successful word segmentation and acquisition of meaning over time. In adults, successful segmentation of pseudowords presented in a continuous syllable stream has been found to be accompanied with an N400 (Cunillera, Toro, Sebastian-Galles, & Rodriguez-Fornells, 2006; Sanders, Newport, & Neville, 2002). Moreover, non-words that resemble real words (i.e., pseudowords) elicit a larger N400 compared to non-words that are not similar to existing words (Braun et al., 2006; Holcomb, Grainger, & O'Rourke, 2002; Rossi, Hartmüller, Vignotto, & Obrig, 2013). During adults' word learning from sentence contexts, N400 responses increased for pseudowords that were presented in a congruent sentence context over time (i.e. a context, that allowed to infer the same meaning across several trials) to match the responses to known real words (Mestres-Misse, Rodriguez-Fornells, & Munte, 2007). However, the N400 responses to pseudowords presented in an inconsistent sentence context did not increase across trials (Mestres-Misse et al., 2007). Together, these effects can be interpreted somewhat opposing to the typical interpretation of the N400, as indicating lexical-semantic integration difficulties or lexical-semantic possibilities. Thus, by recognizing and classifying a phoneme sequence from a pseudoword as a potential real-word, lexical processes are triggered; however, in the case of unsegmented, illegal or infrequent phoneme sequences,

lexical processes are not initiated. In concord with these studies, we interpret the *pairing consistency effect* in our study as indicating an emerging sound-word association in the consistent pairings over time, which are amiss in the case of inconsistent pairings. We argue that this *pairing consistency effect* is a reflection of a learning mechanism which does not necessarily result in the creation of referential relationships between objects and their labels, but clearly indicates associative connections between perceptual representations of objects and pseudowords (c.f. Friedrich & Friederici, 2017; Nazzi & Bertoncini, 2003).

For the testing phase of the current study, we observed the hypothesized *N400-like effect*, such that violated sound-pseudoword pairs evoked significantly more negative responses than matching pairs in the 300-400 ms TW. This is consistent with results of cross-modal object-word mapping studies reporting similar findings for matching versus violated visual object-label combinations (e.g., Borgström et al., 2015b; Friedrich & Friederici, 2008; Junge et al., 2012; Torkildsen et al., 2008; Torkildsen et al., 2007). However, there are differences in the temporal aspects and the topographical distribution of the current *N400-like effect* compared to previous evidence.

Temporally, we observed a left-lateralized *N400-like effect* in an early TW, but only for a short time interval, namely from 300-400 ms. This contrasts previously reported infant N400-like effects, which were longer-lasting, spanning over several 100 ms, and were often found delayed in infants compared to adults (e.g., Junge et al., 2012). Yet, infant N400 effects have been generally reported for various time windows, with an onset as early as 300 ms (Friedrich, Wilhelm, Born, & Friederici, 2015) and even spanning to 1200 ms (Torkildsen et al., 2008). One factor explaining differences in the onset latency of N400-like effects might be age, such that the component occurs earlier with increasing age (for a brief discussion, see Friedrich, Wilhelm, Born, & Friederici, 2015). Given the young age of our tested infant population, however, age might not be the driving factor of the observed early *N400-like effect*. Instead, the exclusive use of the auditory modality in our uni-modal study might explain the early occurrence of the effect, given a documented auditory dominance in infancy (see, Lewkowicz, 1988a, 1988b). Specifically, Robinson and Sloutsky (2010) reported that 10-month-olds when presented with multi-modal stimuli pairs took longer to reach the habituation criterion than infants presented with auditory only stimuli even when visual stimuli were pre-familiarized. As our objects and pseudowords were both presented auditorily, we speculate that the earlier *N400-like effect* could be an indication of faster uni-modal processing. Alternatively, the earlier onset of the *N400-like effect* may be a result of the sequential presentation of the stimuli, such that infants may have been primed to recognize matched pseudowords faster than if the stimuli were presented closer together or in an overlapping manner. Furthermore, the early N400-like effect could also be explained by predictive coding. The N400 has been shown to be influenced by predictability especially in earlier time windows (e.g., Nieuwland et al., 2020). Other experimental designs that allow for the assessment of anticipatory activity in uni- and cross-modal learning paradigms could shed light on this explanation in future experiments.

The topological distribution of our observed *N400-like effect* showed a left-hemispheric dominance. N400 effects have typically been reported to occur over bilateral centro-parietal electrode sites (Kutas & Hillyard, 1980; for a review, see Kutas & Federmeier, 2000). Yet, the left-lateralized distribution in the current study is similar to the N400 distribution reported in an adult study which contrasted environmental sounds that were preceded by related or unrelated words (van Petten & Rheinfelder, 1995). Additionally, an fMRI study conducted by Kiefer, Sim, Herrnberger, Grothe, and Hoenig (2008) reported that words with sound-features (e.g., telephone) elicited higher activation in the left temporal cortex than words without sound-related features. Based on these findings, it is plausible that the topological distribution of the *N400-like effect* found in the current study is influenced by the characteristics of the preceding auditory stimuli, changing the distribution of the *N400-like effect* from a centro-parietal to left-lateralized electrode sites.

Despite differences in temporal and spatial characteristics of the *N400-like effect* in the current study as compared to previous studies, we interpret this ERP effect as evidence of successful association learning during the training phase. An objection to this interpretation might be that infants treat the sound-pseudoword pairs as holistic sound-objects and not as associations between two different auditory stimuli, namely sounds and auditory objects. Studies that have shown that young infants treat object labels in a similar way as other perceptual features of objects (e.g., Sloutsky & Fisher, 2011) suggest that sounds and words could be processed similarly. Yet, while it is possible that sounds and words were processed similarly it is unlikely that they were not discriminated, as it has been shown that even newborns and very young infants discriminate speech from analogous non-speech input (cf. Minagawa-Kawai et al., 2011; Peña et al., 2003).

In summary, in the current study we were able to show that infants are capable of mapping novel labels onto auditory objects in a short time frame with an association-learning paradigm. The replacement of visual objects with auditory objects should not be seen as just a small variation of input. Though it is already known that infants have a preference for the auditory modality (Lewkowicz, 1988a, 1988b) and apply statistical learning to auditory stimuli in sequential presentation (Saffran, 2002; Saffran et al., 1996), it has been unknown whether the auditory modality can serve as an input base for word learning in a similar way as the visual modality. The results of our study provide evidence for the effects of unimodal associative word learning and reveal the modality of the object in object-label pairs as a modulating factor of the N400-like violated expectation effect. This ability is important for linguistic development, as it has been pointed out that the perceptual modality hierarchies are not universal across languages (Majid et al., 2018). In the English language, for example, the auditory modality has been shown to have the second highest number of nouns allocated to it (Winter et al., 2018). Depending on the given language and depending on the input words, auditory semantic features may be important parts of early semantic representations of words and thus important aspects to be learned (cf. Sloutsky & Fisher, 2011; van Petten & Rheinfelder, 1995, for respective findings in adults). The present study indicates that infants may be able to integrate sound-related words into their mental lexicon early on,

even in the absence of any additional visual input, providing evidence for modality-specific representations. Infants might thus be capable of learning the meaning for words like *lullaby*, *siren*, and *thunder* in a similar way to *banana* and *spoon*. Further studies that vary the modality of the object as a modulating factor, while leaving all the other parameters identical, could help to provide additional evidence for modality-specific representations in word learning. Another interesting direction for future research would be the question whether infants are able to build referential links between sounds and words, as, for example, indicated by ERP effects in the testing phase even if sounds and words are presented in the reversed order.

5 Conclusion

In the current study, we were able to show that infants are capable of mapping labels onto auditory objects in a similar way as they do with visual objects. The learning mechanisms triggered in the present study do not necessarily include referential processes, but are best described by domain-general associative learning processes (cf. Sloutsky et al., 2017). The differences between the ERP patterns observed in the present auditory object-pseudoword association study and in previous cross-modal studies need to be investigated in further studies that allow for a direct comparison between different modalities as well as between differences in stimuli presentation, i.e., sequential and simultaneous. Non-visual features are clearly part of (embodied) lexical-semantic representations in adult language processing (Kiefer et al., 2008; Miller, Schmidt, Blankenburg, & Pulvermüller, 2018; Schmidt, Miller, Blankenburg, & Pulvermüller, 2019; Vigliocco et al., 2014). The current study shows how those non-visual features can be acquired and contribute to word learning already before infants' first birthday.

Funding

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - Projektnummer GRK-2185/1 (DFG-Graduiertenkolleg Situated Cognition) (Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number GRK-2185/1 (DFG Research Training Group Situated Cognition)).

Conflict of Interest

The authors have no conflicts of interest to declare.

Acknowledgements

The authors would like to thank lab members from the Department of Neuropsychology at the Max Planck Institute for Human Cognitive and Brain Sciences (MPI-CBS) in Leipzig, Germany and at the Kindersprachlabor at the Institute for Cognitive Science of the University of Osnabrück for their assistance with preparing stimulus material, assisting with experiment programming, contacting

prospective families, and collecting the infant EEG data. Additionally, we would like to thank all of the families who took part in this study.

References

- Banks, M. S., & Salapatek, P. (1983). Infant visual perception. In P. H. Mussen, M. M. Haith, & J. J. Campos (Eds.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 435–571). New York: Wiley.
- Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*(2), 161–169. <https://doi.org/10.1111/j.1469-8986.1993.tb01729.x>
- Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 54–67. <https://doi.org/10.1037/0096-1523.21.1.54>
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*(4), 343–355. [https://doi.org/10.1016/0013-4694\(85\)90008-2](https://doi.org/10.1016/0013-4694(85)90008-2)
- Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), 12916–12921. <https://doi.org/10.1073/pnas.1712966114>
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. <https://doi.org/10.1073/pnas.1113380109>
- Boddy, J. (1986). Event-related potentials in chronometric analysis of primed word recognition with different stimulus onset asynchronies. *Psychophysiology*, *23*(2), 232–245. <https://doi.org/10.1111/j.1469-8986.1986.tb00624.x>
- Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2015a). Event-related potentials during word mapping to object shape predict toddlers' vocabulary size. *Frontiers in Psychology*, *6*(FEB), 1–15. <https://doi.org/10.3389/fpsyg.2015.00143>
- Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2015b). Substantial gains in word learning ability between 20 and 24 months: A longitudinal ERP study. *Brain and Language*, *149*, 33–45. <https://doi.org/10.1016/j.bandl.2015.07.002>
- Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2016). Visual event-related potentials to novel objects predict rapid word learning ability in 20-month-olds. *Developmental Neuropsychology*, *41*(5–8), 308–323. <https://doi.org/10.1080/87565641.2016.1243111>
- Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. <https://doi.org/10.1111/j.0956-7976.2005.01531.x>
- Braun, M., Jacobs, A. M., Hahne, A., Ricker, B., Hofmann, M., & Hutzler, F. (2006). Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. *Brain Research*, *1073–1074*, 431–439. <https://doi.org/10.1016/J.BRAINRES.2005.12.078>
- Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and

- infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160055. <https://doi.org/10.1098/rstb.2016.0055>
- Cummings, A., Čeponiene, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115(1), 92–107. <https://doi.org/10.1016/j.brainres.2006.07.050>
- Cunillera, T., Toro, J. M., Sebastian-Galles, N., & Rodriguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, 1123(1), 168–178. <https://doi.org/10.1016/J.BRAINRES.2006.09.046>
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176. <https://doi.org/10.1126/SCIENCE.7375928>
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9(2), 133–150. [https://doi.org/10.1016/0163-6383\(86\)90025-1](https://doi.org/10.1016/0163-6383(86)90025-1)
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/J.JNEUMETH.2003.10.009>
- Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development*, 5(2–4), 261–276. [https://doi.org/10.1016/S0163-6383\(82\)80036-2](https://doi.org/10.1016/S0163-6383(82)80036-2)
- Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, 266(5604), 718–719. <https://doi.org/10.1038/266718a0>
- Friedrich, M., & Friederici, A. D. (2005). Phonotactic Knowledge and Lexical-Semantic Processing in One-year-olds: Brain Responses to Words and Nonsense Words in Picture Contexts. *Journal of Cognitive Neuroscience*, 17(11), 1785–1802. <https://doi.org/10.1162/089892905774589172>
- Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *Neuroreport*, 19(18), 1757–1761. <https://doi.org/10.1097/Wnr.0b013e328318f014>
- Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: Fast encoding–weak retention. *Journal of Cognitive Neuroscience*, 23(11), 3228–3240. https://doi.org/10.1162/jocn_a_00002
- Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. *Developmental Science*, 20(2). <https://doi.org/10.1111/desc.12357>
- Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications*, 6, 6004. <https://doi.org/10.1038/ncomms7004>
- Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The sleeping infant brain anticipates development. *Current Biology*, 2374–2380. <https://doi.org/10.1016/j.cub.2017.06.070>
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112. <https://doi.org/10.1007/BF02289823>
- Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for Environmental Sound Stimuli. *PLoS ONE*, 8(9), e73382. <https://doi.org/10.1371/journal.pone.0073382>
- Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, 35(1), 66–85. [https://doi.org/10.1016/0093-934X\(88\)90101-0](https://doi.org/10.1016/0093-934X(88)90101-0)
- Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*,

- 14(6), 938–950. <https://doi.org/10.1162/089892902760191153>
- Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of language learning. *Monographs of the Society for Research in Child Development*, 65(3). <https://doi.org/10.1111/1540-5834.00090>
- Horst, J. S., & Samuelson, L. K. (2008). Fast Mapping but Poor Retention by 24-Month-Old Infants. *Infancy*, 13(2), 128–157. <https://doi.org/10.1080/15250000701795598>
- Johnson, E. K. (2016). Constructing a Proto-Lexicon: An Integrative View of Infant Language Development. *Annual Review of Linguistics*, 2(1), 391–412. <https://doi.org/10.1146/annurev-linguistics-011415-040616>
- Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word Form Extraction by 6-month-olds. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0083546>
- Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, 50(14), 3702–3712. <https://doi.org/10.1016/j.neuropsychologia.2012.10.012>
- Junge, C., Cutler, A., & Hagoort, P. (2014). Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy*, 19(2), 179–193. <https://doi.org/10.1111/inf.12040>
- Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, 13(1), 27–39. [https://doi.org/10.1016/S0926-6410\(01\)00085-4](https://doi.org/10.1016/S0926-6410(01)00085-4)
- Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(47), 12224–12230. <https://doi.org/10.1523/JNEUROSCI.3579-08.2008>
- Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., ... Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, 14(3), 220–224. <https://doi.org/10.1111/1467-9280.02435>
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307. <https://doi.org/10.1038/nn1197>
- Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24(1), 109–116. <https://doi.org/https://doi.org/10.1016/j.cogbrainres.2004.12.009>
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463–470. [https://doi.org/10.1016/S1364-6613\(00\)01560-6](https://doi.org/10.1016/S1364-6613(00)01560-6)
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, 207(4427), 203–205. <https://doi.org/10.1126/SCIENCE.7350657>
- Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes*. Hillsdale, NJ: Erlbaum.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the

- N400. *Nature Reviews Neuroscience*, 9(12), 920–933. <https://doi.org/10.1038/nrn2532>
- Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, 24(2), 155–171.
- Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, 24(2), 172–182. <https://doi.org/10.1037/0012-1649.24.2.172>
- Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., ... Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45), 11369–11376. <https://doi.org/10.1073/pnas.1720419115>
- Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, 49(10), 2788–2798. <https://doi.org/10.1016/J.CORTEX.2013.09.003>
- Männel, C., Schaadt, G., Illner, F. K., van der Meer, E., & Friederici, A. D. (2017). Phonological abilities in literacy-impaired children: Brain potentials reveal deficient phoneme discrimination, but intact prosodic processing. *Developmental Cognitive Neuroscience*, 23, 14–25. <https://doi.org/10.1016/J.DCN.2016.11.007>
- Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, 5, 86–94. <https://doi.org/10.1016/J.DCN.2013.01.003>
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4), 831–877. <https://doi.org/10.1037/a0029872>
- Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, 17(8), 1858–1866. <https://doi.org/10.1093/cercor/bhl094>
- Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, 171, 172–179. <https://doi.org/10.1016/J.COGNITION.2017.10.010>
- Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, 5(3), 317–334. <https://doi.org/10.1162/jocn.1993.5.3.317>
- Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, 16(8), 1452–1464. <https://doi.org/10.1162/0898929042304697>
- Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical Brain Imaging Reveals General Auditory and Language-Specific Processing in Early Infant Development. *Cerebral Cortex*, 21(2), 254–261. <https://doi.org/10.1093/cercor/bhq082>
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6(2), 136–142. <https://doi.org/10.1111/1467-7687.00263>
- Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., ... Von Grebmer Zu Wolfsturn, S. (2020). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791), 20180522. <https://doi.org/10.1098/rstb.2018.0522>
- Obrig, H., Mock, J., Stephan, F., Richter, M., Vignotto, M., & Rossi, S. (2017). Impact of associative word learning on phonotactic processing in 6-month-old infants: A combined EEG and fNIRS

- study. *Developmental Cognitive Neuroscience*, 25, 185–197. <https://doi.org/10.1016/J.DCN.2016.09.001>
- Ortu, D., Allan, K., & Donaldson, D. I. (2013). Is the N400 effect a neurophysiological index of associative relationships? *Neuropsychologia*, 51(9), 1742–1748. <https://doi.org/10.1016/j.neuropsychologia.2013.05.003>
- Panneton, R. K., & DeCasper, A. J. (1986). Newborns' postnatal preference for a prenatally experienced melody. In *Paper presented at the International Conference on Infant Studies*. Beverly Hills, CA.
- Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, 23(7), 728–733. <https://doi.org/10.1177/0956797612438734>
- Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumit, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 100(20), 11702–11705. <https://doi.org/10.1073/pnas.1934290100>
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.
- Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Development*, 77(2), 266–280. <https://doi.org/10.1111/j.1467-8624.2006.00869.x>
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29(6), 819–865. https://doi.org/10.1207/s15516709cog0000_31
- Rhodes, S. M., & Donaldson, D. I. (2007). Association and not semantic relationships elicit the N400 effect: Electrophysiological evidence from an explicit language comprehension task. *Psychophysiology*, 0(0), 070914092401002-???. <https://doi.org/10.1111/j.1469-8986.2007.00598.x>
- Richards, D. D., & Goldfarb, J. (1986). The episodic memory model of conceptual development: An integrative viewpoint. *Cognitive Development*, 1(3), 183–219. [https://doi.org/10.1016/S0885-2014\(86\)80001-6](https://doi.org/10.1016/S0885-2014(86)80001-6)
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, 75(5), 1387–1401. <https://doi.org/10.1111/j.1467-8624.2004.00747.x>
- Robinson, C. W., & Sloutsky, V. M. (2010). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology*, 107(3), 351–358. <https://doi.org/10.1016/j.jecp.2010.04.006>
- Rossi, S., Hartmüller, T., Vignotto, M., & Obrig, H. (2013). Electrophysiological evidence for modulation of lexical processing after repetitive exposure to foreign phonotactic rules. *Brain and Language*, 127(3), 404–414. <https://doi.org/10.1016/J.BANDL.2013.02.009>
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47(1), 172–196. <https://doi.org/10.1006/JMLA.2001.2839>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, 69(1), 94–104. <https://doi.org/10.1111/j.1467-8624.1998.tb06136.x>
- Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential

- index of perceived onsets in continuous speech. *Nature Neuroscience*, 5(7), 700–703. <https://doi.org/10.1038/nn873>
- Schmidt, T. T., Miller, T. M., Blankenburg, F., & Pulvermüller, F. (2019). Neuronal correlates of label facilitated tactile perception. *Scientific Reports*, 9(1), 1606. <https://doi.org/10.1038/s41598-018-37877-w>
- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), 6038–6043. <https://doi.org/10.1073/pnas.1017617108>
- Sloutsky, V. M., & Fisher, A. V. (2011). *The Development of Categorization. Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 54). <https://doi.org/10.1016/B978-0-12-385527-5.00005-X>
- Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, 74(3), 822–833. <https://doi.org/10.1111/1467-8624.00570>
- Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology*, 97, 1–30. <https://doi.org/10.1016/J.COGLPSYCH.2017.06.001>
- Smith, L. B., Jayaraman, S., Clerkin, E. M., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 325–336. <https://doi.org/10.1016/J.TICS.2018.02.004>
- Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition*, 60(2), 143–171. [https://doi.org/10.1016/0010-0277\(96\)00709-3](https://doi.org/10.1016/0010-0277(96)00709-3)
- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558–1568. <https://doi.org/10.1016/j.cognition.2007.06.010>
- Tabullo, A., Yorio, A., Zanutto, S., & Wainelboim, A. (2015). ERP correlates of priming in language and stimulus equivalence: Evidence of similar N400 effects in absence of semantic content. *International Journal of Psychophysiology*, 96(2), 74–83. <https://doi.org/10.1016/j.ijpsycho.2015.03.004>
- Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2019). The Limits of Infants' Early Word Learning. *Language Learning and Development*, 1–21. <https://doi.org/10.1080/15475441.2019.1670184>
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10(1), 21. <https://doi.org/10.1186/1471-2202-10-21>
- Teixidó, M., François, C., Bosch, L., & Männel, C. (2019). The role of prosody in early speech segmentation and word-referent mapping: Electrophysiological evidence. In P. P. Fabra & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition*. (pp. 79–100). Amsterdam: John Benjamins.
- Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *NeuroReport*, 14(18). <https://doi.org/10.1097/00001756-200312190-00004>
- Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds, 10(2), 172–175. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1111/1467-9280.00127>
- Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432–444. <https://doi.org/10.1111/j.1532-7078.2011.00084.x>
- Torkildsen, J. von K., Friis Hansen, H., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of

productive vocabulary size. *Brain and Language*, 108(2), 73–88.
<https://doi.org/10.1016/J.BANDL.2008.09.005>

Torkildsen, J. von K., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience*, 20(7), 1266–1282.
<https://doi.org/10.1162/jocn.2008.20087>

Torkildsen, J. von K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, 20(4), 332–351. <https://doi.org/10.1016/J.JNEUROLING.2007.02.003>

van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33(4), 485–508.
[https://doi.org/10.1016/0028-3932\(94\)00133-A](https://doi.org/10.1016/0028-3932(94)00133-A)

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, 24(7), 1767–1777. <https://doi.org/10.1093/cercor/bht025>

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, 34(6), 1289–1309.
<https://doi.org/10.1037/0012-1649.34.6.1289>

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179, 213–220.
<https://doi.org/10.1016/j.cognition.2018.05.008>

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/J.COGNITION.2012.06.016>

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective Attention in Cross-Situational Statistical Learning: Evidence From Eye Tracking. *Frontiers in Psychology*, 3.
<https://doi.org/10.3389/fpsyg.2012.00148>

A)		B)					
Consistent Pairs		Inconsistent Pairs		Matching Pairs		Violated Pairs	
1st - 4th	Sound 1 Pseudoword 1	Sound 9 Pseudoword 9	Sound 1 Pseudoword 1	Sound 1 Pseudoword 3			
	Sound 1 Pseudoword 1	Sound 10 Pseudoword 9	Sound 2 Pseudoword 2	Sound 2 Pseudoword 8			
	Sound 1 Pseudoword 1	Sound 11 Pseudoword 9	Sound 3 Pseudoword 3	Sound 3 Pseudoword 2			
	Sound 1 Pseudoword 1	Sound 12 Pseudoword 9	Sound 4 Pseudoword 4	Sound 4 Pseudoword 1			
5th - 8th	Sound 1 Pseudoword 1	Sound 13 Pseudoword 9	Sound 5 Pseudoword 5	Sound 5 Pseudoword 7			
	Sound 1 Pseudoword 1	Sound 14 Pseudoword 9	Sound 6 Pseudoword 6	Sound 6 Pseudoword 5			
	Sound 1 Pseudoword 1	Sound 15 Pseudoword 9	Sound 7 Pseudoword 7	Sound 7 Pseudoword 4			
	Sound 1 Pseudoword 1	Sound 16 Pseudoword 9	Sound 8 Pseudoword 8	Sound 8 Pseudoword 6			

C)

Trial =	Event	Sound	Pause	Pseudoword	Inter-trial Pause
	Duration	950 ms	600 ms	750 ms	1600 ms

Figure 1: Experimental Design: (A) gives an example of how the pseudowords are distributed in the two conditions of the training session. (B) Example of the distribution of pseudowords in the testing phase. (C) The construction of all trials in the experiment, both for the training and testing phases.

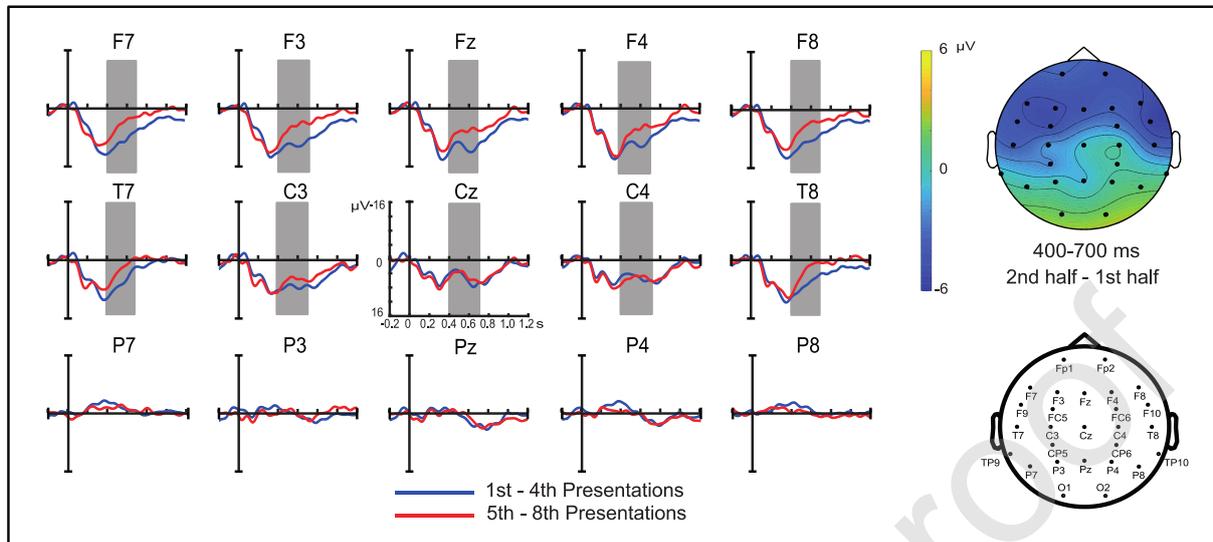


Figure 2: Word-form Familiarity: The ERPs time-locked to the onset of the pseudoword in the training phase (10 Hz low-pass filter applied for visualization only). Blue lines indicate the first-fourth presentations of all pseudowords and the red lines indicate the fifth-eighth presentation of all pseudowords. The scalp maps depict the spatial distribution of the difference in the ERP amplitude between the 1st through 4th presentations and the 5th through eighth presentations in the given window.

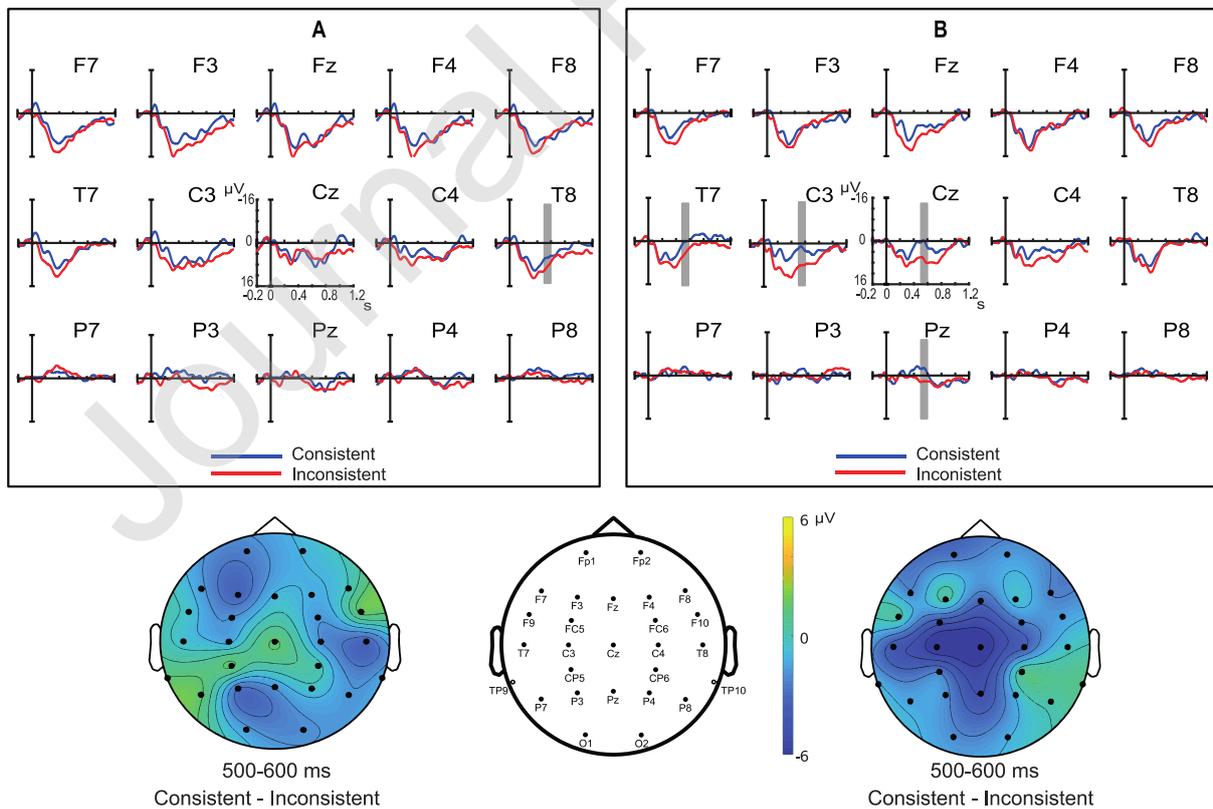


Figure 3: Pairing Consistency Effect: (A) ERPs time-locked to the onset of consistent and inconsistent pseudowords in the first half of the training phase; (B) ERPs time-locked to the onset of consistent and inconsistent pseudowords in the second half of the training phase (10 Hz low-pass filter applied for visualization only). Blue lines indicate the consistent pairings and the red lines indicate the inconsistent pairings. The scalp maps depict the spatial distribution of the difference in the ERP amplitude between consistent and inconsistent words in the given time window.

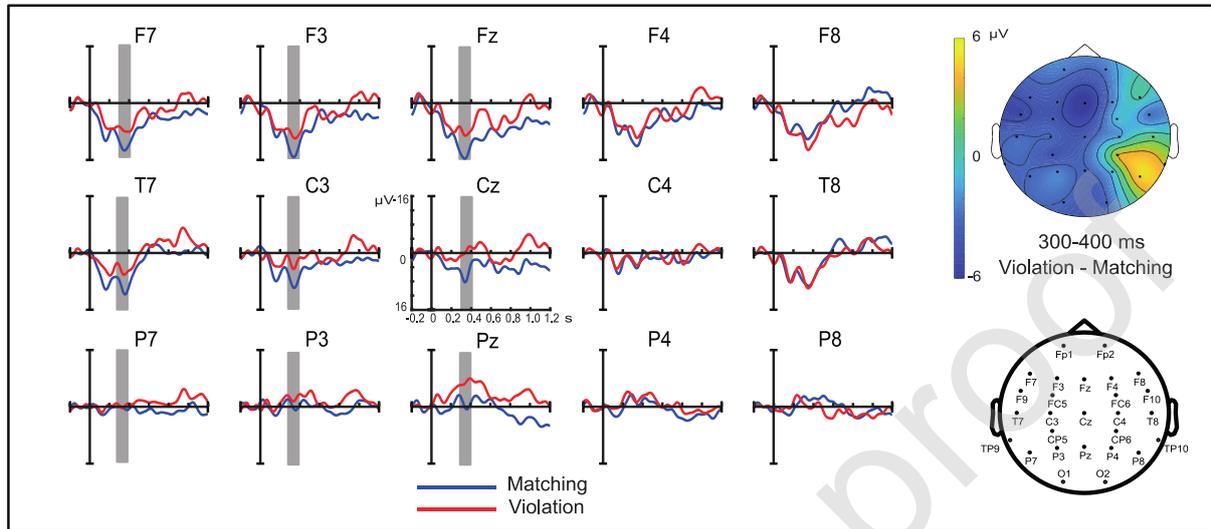


Figure 4: Matching vs Violation: The ERPs time-locked to the onset of the pseudoword in the testing phase (10 Hz low-pass filter applied for visualization only). Blue lines indicate the matching object-word pairs and the red lines indicate the violated object-word pairs. The scalp maps depict the spatial distribution of the difference in the ERP amplitude between matching and violated words in the given time window.

Table 1: Step-Down Analysis of Word-form Familiarity Effect: Statistic analysis of ANOVA for significant regions and step-down analysis for word-form familiarity effect in the training phase

Time Window (in ms)	Repetition x Region (ANOVA)	Step-Down Analysis of Repetition (<i>t</i> -test)
400-500	$F(1.33,41.31)=7.149, p = 0.006^{\circ}$	anterior: $T(31) = 2.7722, p = 0.0093^*$ central: $T(31) = 3.0811, p = 0.0043^*$
500-600	$F(1.47,45.55)=7.385, p=0.004^*$	anterior: $T(31) = 3.4689, p = 0.0016^{**}$ central: $T(31) = 3.3613, p = 0.0021^{**}$
600-700	$F(1.48,45.76)=8.740, p = 0.002^*$	anterior: $T(31) = 3.9156, p = 0.0005^{**}$

central: $T(31) = 2.9335, p = 0.0063^*$

$^{\circ}p < 0.1$ $*p < 0.05$ $**p < 0.01$ $***p < 0.001$

Journal Pre-proof