

Auditory associative word learning in adults: the effects of musical experience and stimulus ordering

Samuel H. Cospers^{1*}, Claudia Männel^{2,3}, & Jutta L. Mueller⁴

¹Chair of Lifespan Developmental Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

²Department of Audiology and Phoniatics, Charité–Universitätsmedizin Berlin, Berlin, Germany

³Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

⁴Department of Linguistics, University of Vienna, Vienna, Austria

*Corresponding author at: Technische Universität Dresden, Chair of Lifespan Developmental Neuroscience, Faculty of Psychology, 01062 Dresden, Germany

Abstract:

Evidence for sequential associative word learning in the auditory domain has been identified in infants, but adults have proven to have difficulties in this type of word learning. To investigate the role of auditory expertise and auditory stimuli salience in the association of auditory objects and their labels, we tested in the first experiment auditorily-trained musicians versus athletes as a high-level control group and in the second experiment stimulus ordering, contrasting object-label versus label-object presentation. Learning was evaluated from electrophysiological (EEG) recordings during training and subsequent testing phases, as well as accuracy-judgement responses during test. Results revealed for musicians a late positive component in the EEG during testing, neither N400 nor behavioral effects were found at test, while athletes did not show any effect of learning. Moreover, the object-label-ordering group only exhibited emerging association effects during training, while the label-object-ordering group showed an EEG effect during test as well as above chance accuracy-judgement scores. Thus, our results suggest an advantage of musical expertise and more salient label-object-ordering in auditory associative word learning in adults. Implications with respect to developmental differences in auditory word learning are discussed.

Keywords: associative word learning, ERP, N400, LPC, musicians, auditory modality

1 Introduction

To what extent is word learning influenced by perceptual modality? Traditionally, word learning experiments have focused on labeling visual objects (e.g., Breitenstein et al., 2005; Friedrich & Friederici, 2008, 2011; Horst & Samuelson, 2008; McMurray et al., 2012; Sloutsky et al., 2017; Taxitari et al., 2019; Werker et al., 1998). However, recent research has expanded into word learning for entities from other modalities such as tactile (e.g., Miller et al., 2018; Schmidt et al., 2019), olfactory (e.g., Vanek et al., 2021), gustatory (e.g., Yan et al., 2021), and auditory (Casper et al., 2020, 2022). This distribution of word-learning research in modality does reflect the distribution of sensory modality attributes in the vocabularies of many languages, where the largest attribution of words is to the visual modality (e.g., Chedid et al., 2019; Chen et al., 2019; Lynott et al., 2020; Miklashevsky, 2018; Speed & Brybaert, 2022; Vergallito et al., 2020). In contrast, the auditory modality occupies a middle position in this distribution in most languages, although there is no universal hierarchy of perceptual modality across different vocabularies (Majid et al., 2018). Regarding auditory word learning, it has been shown in early development that words for environmental sounds can be learned associatively in a similar manner compared to words for visual objects (Casper et al., 2020). Thus, infants are able to build associative links between a word (or pseudoword) and an auditory object (e.g., an environmental sound) over time (e.g., McMurray et al., 2012; Regier, 2005; Sloutsky et al., 2017). While adults can also utilize other mechanisms to achieve word learning, such as inferential learning by means of hypothesis testing (cf. Sloutsky et al., 2017; Xu & Tenenbaum, 2007), associative learning remains vital to word learning across the lifespan (Dittinger et al., 2016, 2019; Regier, 2005; Sloutsky et al., 2017). First evidence of associative auditory word learning for adults, however, suggested that their learning of words for sounds is more limited than in children (Casper et al., 2020, 2022). How strong this limitation is and by what learning factors it might be overcome is yet to be investigated.

The present word-learning study in adults asked whether the above described modality-related constraint on learning can be overcome by external and internal learning factors. Specifically, we aimed to investigate whether behavioral and electrophysiological (EEG) correlates of auditory object–label association can be modulated by i) expertise in the auditory domain and by ii) enhancing the saliency of associated stimuli through the order of presentation. Expertise in the auditory domain will be operationalized by musicianship and contrasted against a high-level control group of athletes. Saliency of the associated stimuli will be operationalized by reversing the order of stimulus presentation compared to our previous word-learning studies

with object-label ordering in infants and adults (Casper et al., 2020, 2022), such that the more salient auditory labels will be presented first across learning trials.

In terms of the cognitive benefits of musical training, musicians have been generally shown to outperform non-musicians in visual, phonological, and executive memory tasks as well as a faster updating of working memory (George & Coch, 2011). Regarding expertise in the auditory modality, musicians have been shown to possess a superior auditory short-term memory as compared to non-musicians (Chan et al., 1998; Cohen et al., 2011; Ho et al., 2003). In an analysis of auditory statistical learning, musicians also showed superior abilities (Pesnot Lerousseau & Schön, 2021) and exhibited enhanced global efficiency and density of the cortical network involved in the learning of tonal sequences as compared to non-musicians (Paraskevopoulos et al., 2017). Moreover, there is evidence that musical expertise and its enhanced cognitive attributes may even transfer to language learning (Schön & Francois, 2011). Both adult musicians and musically-trained children showed improved word learning (visual objects labeled with spoken pseudowords) as compared to non-musicians (Dittinger et al., 2016, 2017).

The decision of including athletes as a control group in the current study is based on the fact that they are also a high-expertise population, yet unrelated to auditory skills (for a similar procedure, see Dittinger et al., 2016). Interestingly, athletes demonstrate positive influences of embodied sport practices in a range of various cognitive tasks, such as visuo-spatial attention processing, and higher level cognitive tasks, such as executive control (Alves et al., 2013). Additionally, athletes exhibit advantages in attentional cuing, processing speed, and varied attention paradigms as compared to non-athletes or non-professional athletes (Voss et al., 2010). Implicit memory advantages have also been found for experienced athletes, although these results are limited to the context of the sport (e.g., soccer players in soccer-specific situations; Zoudji & Thon, 2003). Furthermore, elite-athletes exhibited an advantage over recreational-athletes in sustained visual attention, but again these advantages are not extended to non-sports-specific cognitive measures, such as verbal memory span (Heppe et al., 2016).

In addition to auditory expertise, we also investigated the effect of stimulus saliency by manipulating the order in which stimuli are presented. This is particularly relevant for auditory associative word learning, as auditory information is by nature fleeting and does not persist over time. While environmental sounds and speech are processed in similar ways in principle, for example, at the level of spectrotemporal analysis (Klein et al., 2003), but also with respect to processing of context and meaning (Ballas & Howard, 1987; van Petten & Riefelder,

1995), speech is more salient. As indicated by activation patterns in the auditory cortex and in behavioral analyses, discrimination of speech is better than discrimination of environmental sounds in natural auditory scenes containing both stimuli (Renvall et al., 2021). However, neural effects of conceptual priming (N400) have indicated that familiar environmental sounds may elicit an earlier onset of the N400 than those elicited by words, although conceptual processing of verbal and non-verbal stimuli is considered to be similar in nature (Orgs et al., 2006). Behavioral measures have also shown increased accuracy rates in the processing of auditory-verbal information in relation to auditory-nonverbal information both in the processing of visual objects pairings with either linguistic or non-linguistic stimuli (Calignano et al., 2021) and in statistical learning (Lukics & Lukács, 2022; Thiessen, 2010). Short-term memory accuracy is also higher in tasks with verbal stimuli than with nonverbal stimuli in both the auditory and the visual modalities (Talamini et al., 2022). Taken together, these studies suggest a processing and memory advantage for speech compared to non-speech stimuli. Yet, when sequences of speech and non-speech stimuli are presented together, not only does the saliency of each single stimulus play a role, but also how the two interact with each other. We know that both words and environmental sounds can be used as cues to meaning in principle (Orgs et al., 2006; van Petten & Rheinfelder, 1995), but we do not know how the factors detailed above, saliency and memorizability, may impact on learning if auditory objects and spoken labels are presented in sequence.

The current study uses an adaptation of the word learning paradigm of Friedrich and Friederici (2008, 2011), with a combination of behavioral and electrophysiological measurements to allow for an in depth analysis of associative word learning in adults. In two experiments, auditory objects (environmental sounds) and pseudowords (labels) are presented sequentially in pairs across learning and testing phases, while an EEG is recorded and accuracy ratings are obtained. First, two expert populations are tested, musicians and athletes, and second, a general-population group is tested in a label-object-ordering experiment and compared to a previously reported study using object-label ordering (Casper et al. 2022). For the musicians, we predicted the group would show behavioral (i.e., above-chance accuracy on a judgement task) and electrophysiological evidence of word learning in the testing phase (i.e., N400 effect). For the athletes, we predicted this group would be similar to the auditory-sequential condition of Casper and colleagues (2022) and not show learning effects. For the second experiment, with reversed label-object ordering, we expected learning effects due to an attentional and memory advantage of labels that should facilitate linking related stimuli in the learning task, as evidenced in above-chance accuracy in the judgment task as well as an N400 effect in the EEG.

2 Experiment 1: Musicians and Athletes

2.1 Materials and methods

2.1.1 Participants

A total of 48 participants (24 females and 24 males) with ages ranging from 19 to 29 years of age ($M = 23.32$ years; $SD = 2.84$ years) in two different groups participated in the study: musicians and athletes. All participants were found through the University of Osnabrück, native German speakers, right-handed, reported normal hearing and normal/corrected eye-sight, and did not suffer from any neurological conditions (including learning impairments, ADHD, or dyslexia). In order to be included in the final dataset, each participant must have retained at least 75% of each trial type after artifact rejection and ICA correction, which is described in more detail below.

For the musicians, a group of 24 students participated, one participant was not a student but had 10 years of experience playing in a professional symphony. One participant ($n=1$) was excluded, as inclusion criteria after pre-processing was not met. The final dataset for musicians included 23 individuals (11 female), aged 19 to 29 ($M = 24.02$ years; $SD = 5.57$ years). In order to be included in the study, musicians had to have been playing an instrument for at least 11 years (for similar requirements, see Dittinger et al., 2016) as well as being enrolled as a music student in a Bachelor or Master program at the University of Osnabrück (see above for the one exception). The musicians reported having played an instrument for an average of 15.04 years ($SD = 2.59$ years). Instruments reported included 10 pianos, seven violins, two trumpets, one drums, one horn, one cello, and one trombone.

For the athletes, 24 students (12 female) participated with ages ranging from 19 to 29 years of age ($M = 22.74$ years; $SD = 2.68$ years). In order to be included in the study, athletes needed to be active in sports with regular training sessions (at least 3.5 hours a week) for at least the past five years. Athletes reported participating in competition sports ranging between 5 to 19 years ($M = 10.33$ years; $SD = 3.59$ years), and training from 3.5 to 7 hours a week ($M = 4.84$ hours per week; $SD = 1.01$ hours). Sport disciplines were reported in groups: team sports (soccer/football, handball, volleyball, water polo, hockey, and table tennis; $n = 14$), individual sports (swimming, track and field, triathlon, bouldering; $n = 7$), and martial arts (judo, kickboxing, and karate; $n = 3$). For an overview of all participants, see Table 1.

None of the participants of this experiment had previously participated in other auditory associative word learning experiments. Participants of all groups gave written informed consent

before participation. All participants were compensated for their time with either credit towards their degree program (mandatory participation in experiments for Psychology and Cognitive Science Bachelor programs) or 10 Euro. The experiment was approved by the ethics committee of the University of Osnabrück and conformed to all aspects of the Declaration of Helsinki (World Medical Association, 2013).

Participant Group	Sample Size	Average Age (SD)	Average years trained (SD)	1 st Half Training – Consistent (SD)	1 st Half Training – Inconsistent (SD)	2 nd Half Training – Consistent (SD)	2 nd Half Training – Inconsistent (SD)	Testing – Matching (SD)	Testing – Violated (SD)
Musicians	23	24.02 years (5.57 years)	15.04 years (2.59 years)	30.8 trials (1.5 trials)	30.7 trials (1.9 trials)	30.9 trials (1.3 trials)	30.6 trials (1.5 trials)	14.9 trials (1.1 trials)	14.8 trials (1 trial)
Athletes	24	22.74 years (2.68 years)	10.33 years (3.59 years)	30.3 trials (1.9 trials)	30.2 trials (1.9 trials)	29 trials (3.1 trials)	28.5 trials (4.6 trials)	13.8 trials (2.5 trials)	14 trials (2.2 trials)
Object-Label-Ordering*	22	23.86 years (3.4 years)	-	31.2 trials (1.2 trials)	31.2 trials (1.2 trials)	30.4 trials (1.6 trials)	30.6 trials (0.7 trials)	15.5 trials (0.7 trials)	15.1 trials (1 trial)
Label-Object-Ordering	23	23.36 years (5.23 years)	-	29.9 trials (3.1 trials)	30.1 trials (2.2 trials)	29.2 trials (3.1 trials)	29.6 trials (2.3 trials)	13.7 trials (2.2 trials)	14.1 trials (1.9 trials)

Table 1: Overview of participants by sample size and age per group along with average time played (instrument for musicians, competition sports for athletes, and non-applicable for the object-label-ordering group and the label-object-ordering group) and average number of trials per condition included in the final dataset per participant group. Standard deviation was also indicated for all averages. The participant information from the object-label-ordering group (marked *) was from a previous publication (auditory sequential; Cospers et al., 2022).

2.1.2 Stimuli

The stimuli utilized in the current experiment contained 16 environmental sounds from the NESSTI database for environmental sounds (Hocking et al., 2013), cut to 950 ms in length, and 16 pseudowords derived from existing German nouns with sound files 750 ms in length. All stimuli were previously used in auditory associative word learning experiments in both infants and adults (Cospers et al., 2020, 2022). All files were saved as monaural .WAV files, digitized at 44100 Hz, and presented over two speaker channels at a comfortable volume. The experiment

was presented using the Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA, USA, version 20.0).

2.1.3 Procedure

For the experiment, participants were seated in front of a computer monitor, outfitted with speakers on either side. Buttons for the behavioral analysis were on the table within reach. Participants were instructed to pay attention to the sounds and words presented, as they would be asked questions about them at a later point, while looking at the screen. Ultimately, the participants were ignorant to the goal of the study and were not explicitly instructed to attend the sound-pseudoword (or pseudoword-sound) pairings and their consistency. The experiment was split into two sections: a training phase and a testing phase. The training phase consisted of 128 trials in which environmental sounds and pseudoword combinations were presented. Of the 16 environmental sounds, eight were consistently presented with a single pseudoword eight times each (for a total of 64 trials). The remaining eight environmental sounds and eight pseudowords were paired with one another only once, creating 64 unique combinations in as many trials (for similar training phases in associative word learning, see Cosper et al., 2020, 2022; Friedrich & Friederici, 2008, 2011). Inconsistent pairings were included to test online association during the learning phase. The testing phase utilized only the environmental sounds and pseudowords of the consistent pairings in the training phase, presented either matching (identical to the consistent pairing in training) or violated (new combinations of sounds and pseudowords, violating built pairings in training). Each pairing was presented twice for a total of 16 pairings in each condition and 32 trials overall. Figure 1A provides an example of the sound-pseudoword combinations in the training phase and Figure 1B gives an example of the testing phase. Eight randomized lists were created to counterbalance stimuli across training conditions, redistribute pairings, and to invert presentation lists, in order to test overall associative word learning abilities and not individual stimuli pairing preferences. A fixation cross was always presented on the screen.

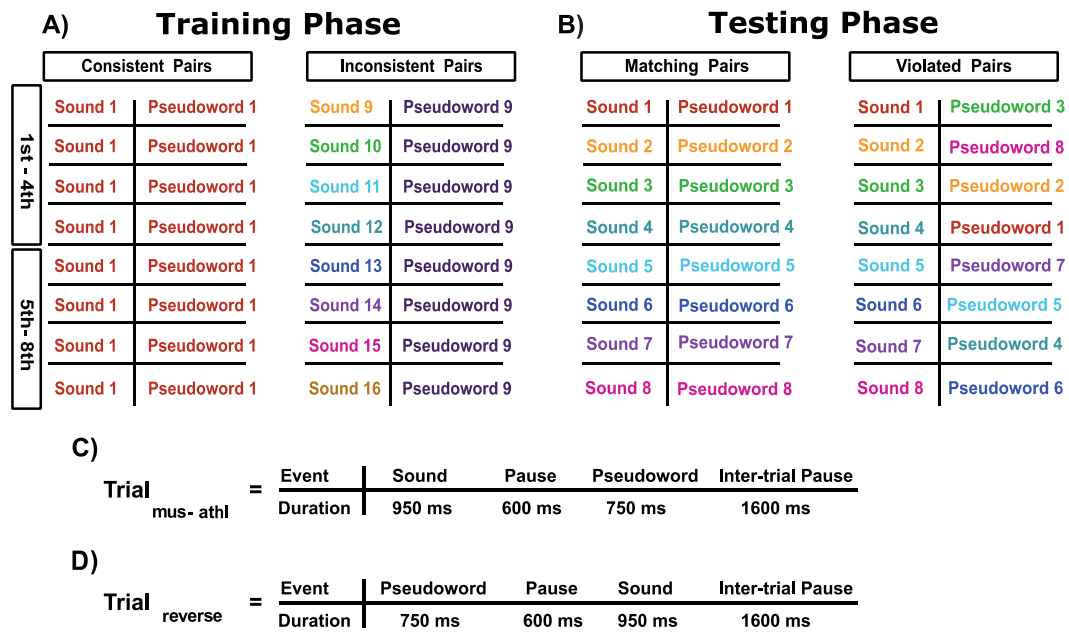


Figure 1: Experimental setup. (A) Example of pairings of sound-pseudoword combinations for the training phase as presented to the musician and athlete groups. (B) Example of pairings of sound-pseudoword combinations for the testing phase as presented to the musician and athlete groups. (C) Trial structure for the musician and athlete groups. (D) Trial structure for the label-object-ordering group. Colors used for visual emphasis. Figure adapted from Cosper and colleagues (2022).

Each trial was structured identically across the experiment. For experiment 1, sounds were presented first, then a 600 ms inter-stimulus pause, followed by the pseudoword (see Figure 1C). After the final stimulus in the pairing was presented, a 1600 ms inter-trial pause was presented. Between the training and testing phase, participants has a 1-2-minute break, in which the instructions for the testing phase were also given. In the testing phase, participants were additionally asked to provide behavioral input after the presentation of each pairing, indicating if the pairing was correct by means of button pressing. The assignment of left and right buttons (for yes and no answers) was counterbalanced across the randomization lists. The experiment lasted 14 minutes in total.

2.1.4 EEG Processing

All data were collected at the Kindersprachlabor of the University of Osnabrück. EEG signals were measured with a REFA amplifier (Twente Medical Systems International, Oldenzaal, The Netherlands) using Ag/AgCl electrodes at extended standard 10-20 electrode positions in 64-

channel TMSi caps (Twente Medical Systems International, Oldenzaal, The Netherlands). Eye movement-related activities were measured by monopolar electrodes placed above and below the left eye and on either temple. Impedances were kept below 5 k Ω . The EEG data were recorded continuously using the TMSi Polybench software (Twente Medical Systems International, Oldenzaal, The Netherlands) at a sampling rate of 512 Hz, average reference, and a ground electrode located on the left collarbone.

Pre-processing of EEG data was done in MATLAB (The Mathworks Inc., Natick, MA, USA, version 2022b) using EEGLAB (Delorme & Makeig, 2004, version 14_1_1b) following the same protocol as Cospers and colleagues (2022). First, a high-pass filter of 1 Hz (-3dB, cutoff frequency of 1.38 Hz) and a low-pass filter of 30 Hz (-3dB, cutoff frequency of 31.16 Hz) were applied to the continuous EEG data. The data was then epochized, with epochs time-locked to the onset of the second stimulus in the presented pairing with a length of 1200 ms and a pre-stimulus baseline of 200 ms. For clarity, the analyses for musician, athlete, and object-label-ordering groups were conducted on the pseudoword and the analyses for the label-object-ordering group were conducted on the environmental sound. Subsequently, EEG data were re-referenced to linked mastoids. Artifact rejection was conducted by means of manual inspection. Following, a semi-automatic independent component analysis (ICA) was applied to each dataset in order to correct eye-movement artifacts. Afterwards, the ICA weights of the 1 – 30 Hz filtered dataset were applied to a dataset with a high-pass filter of 0.3 Hz (-3dB, cutoff frequency of 0.36 Hz) and a low-pass filter of 30 Hz (-3dB, cutoff frequency of 31.16 Hz). All statistical analyses were conducted using this 0.3 – 30 Hz dataset. As mentioned above, participants must have met the inclusion criteria to remain in the final dataset. The average number of trials and their standard deviations for each participant group can be seen in Table 1.

2.1.5 Statistical Analysis

The behavioral results are presented as the accuracy of participants' responses and was assessed in R (R Core Team, 2020) using a one-sample t-test against chance, $\mu = 0.5$.

EEG data were statistically analyzed for each participant group in MATLAB (The Mathworks Inc., Natick, MA, USA) using the FieldTrip toolbox (Oostenveld et al., 2011, version 20220104). Cluster-based permutation analyses were conducted by means of dependent samples t-tests for each sampling point with a set alpha threshold of 0.05. A minimum number of neighborhood channels was set to 2 for a significant sample to be included in the clustering

algorithm. A permutation test, using the Monte Carlo method (Maris & Oostenveld, 2007), was applied with 1000 randomizations and an alpha threshold of 0.05. Reference electrodes and EOG electrodes were excluded from the cluster-based permutation test. Two time windows (TWs) were selected for the analyses. An N400 TW was selected from 400 – 800 ms post-stimulus, pre-selected in order to compare to previous auditory associative word learning experiments (Cosper et al., 2022). Additionally, a later TW was selected to assess possible later effects: 800 – 1200 ms.

The EEG statistical analyses were conducted in three comparisons: first half of the training phase, second half of the training phase, and the testing phase. For the first half of the training phase, cluster-based permutation tests were applied to test differences between consistent pairings versus inconsistent pairings for the first through fourth presentations of each target stimulus (pseudoword or environmental sound). For the second half of the training phase, cluster-based permutation tests were applied to the fifth through eighth presentations of each target stimulus to assess differences between consistent and inconsistent pairings. Violated and matching pairings were assessed in the testing phase for each group. It is important to note that as cluster-based permutation tests are applied here, TWs reported in the results section are not meant to be determinations of absolute TWs of effects nor are cluster-based permutation tests suitable for the determination of effect latency (cf. Sassenhagen & Draschkow, 2019). Thus, results presented below are approximate effect TWs and only describe clusters of electrodes in which significant differences are found.

2.2 Results

2.2.1 Behavioral Results

The behavioral analysis of the musician group showed that participants correctly identified correct and incorrect sound-pseudowords pairings in the testing phase only 52% of the time ($M = 16.74$ correct responses; $SD = 2.59$ responses). A one-sample t-test indicated that the musicians were not above chance identifying matching and violated pairings, $t(22) = 1.3379$, $p = 0.195$ (see Figure 2).

The behavioral analysis of the athletes group yielded a 56% accuracy ratio ($M = 17.96$ correct responses; $SD = 5.95$ responses). A one-sample t-test indicated that athletes were not above chance in correctly identifying matching and violated sound-pseudoword pairings in the testing phase, $t(23) = 1.5791$, $p = 0.128$ (see Figure 2).

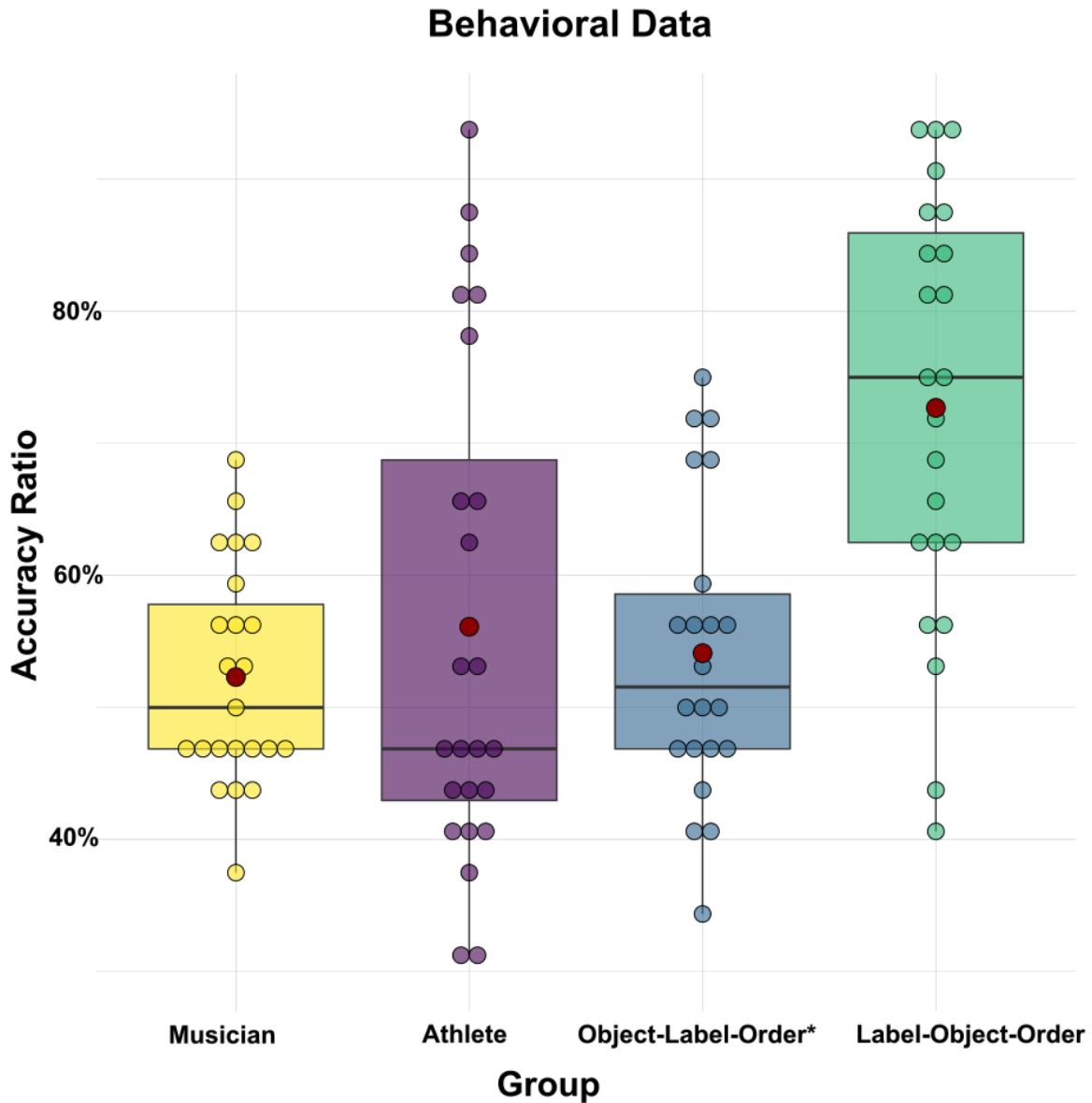


Figure 2: Behavioral results of the testing phase. The distribution of the accuracy ratios for each participant is presented per group. Chance level is at 50%. The boxes represent the upper and lower quartile, while the whiskers represent variability beyond the quartiles. The line within the boxes represents the median accuracy ratio, the red dot represents the mean, and the colored dots represent the individual accuracy ratios of the participants. The data for the Object-Label-Ordering group is from a previous publication (auditory sequential; Cosper et al., 2022).

2.2.2 EEG: Training Phase

For the musician group, no clusters with significant processing differences between inconsistent over consistent sound-pseudoword pairings in the first or the second half of training were found in either the 400 – 800 ms or the 800 – 1200 ms TW (see Figure 3A).

In the athlete group, no clusters showing significant processing differences between inconsistent and consistent sound-pseudoword pairings were found in the first or the second half of the training phase in either the N400 or late TW (see Figure 3B).

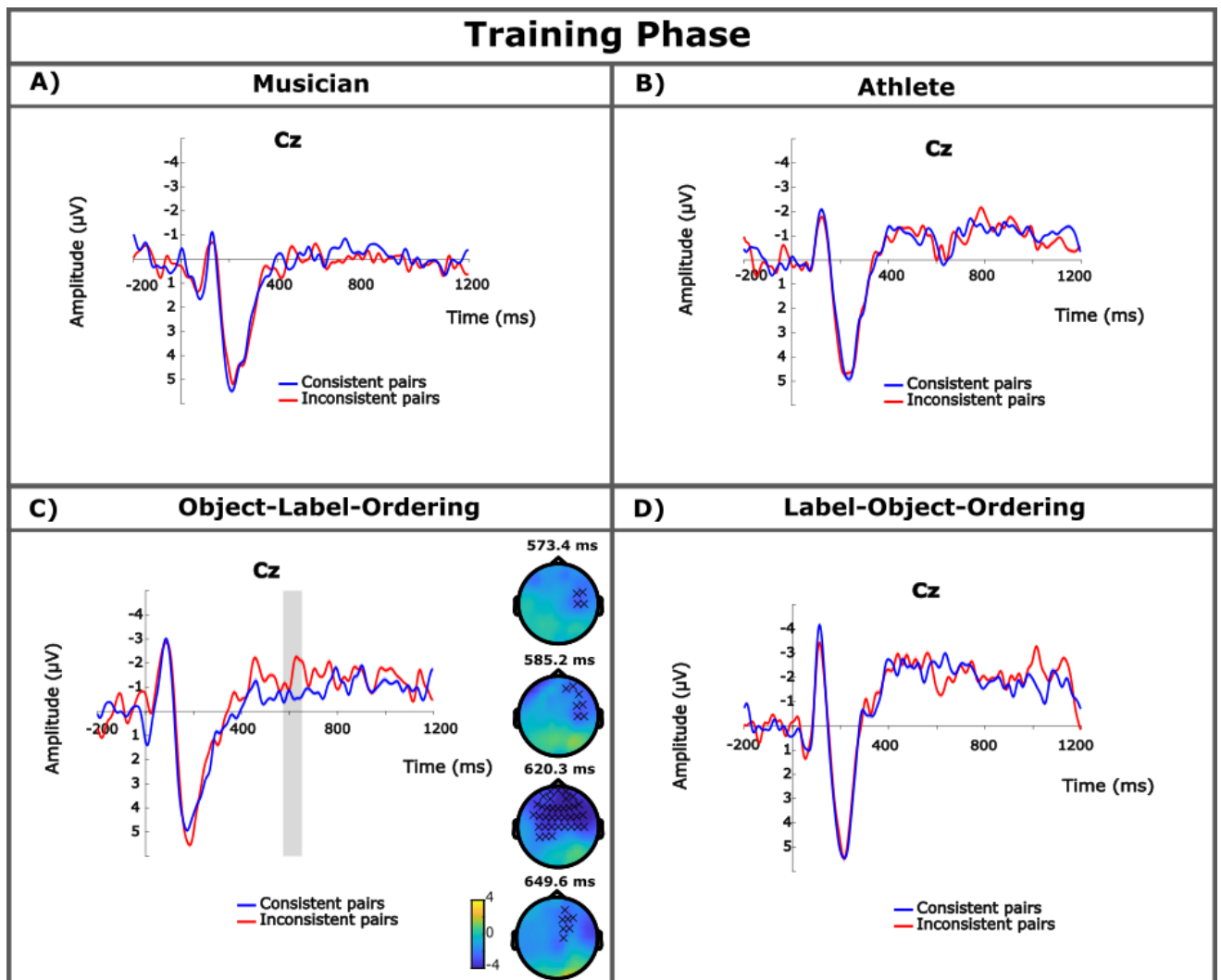


Figure 3: Training phase results. ERPs are presented time-locked to the onset of the second stimuli in the pairing for the second half of the training phase of each experiment (A, B, C represent pseudowords, D represents environmental sounds). The Cz electrode was selected as a representative electrode for all of the experiments. The blue lines indicate the consistent pairs while the red lines indicate the inconsistent pairs. The lighter-shaded lines indicate the standard

error of the waveform. Gray bars are used to illustrate the time window where significant differences were found between the conditions. Topological difference maps from the cluster-based permutation test are presented only where significant differences were found and show the electrode distribution in the cluster over time (the X symbol indicates $p < 0.05$). Results of Object-Label-Ordering (C) are reprinted from a previous publication (Cospers et al., 2022).

2.2.3 EEG: Testing Phase

The cluster-based permutation in the testing phase of the musician group yielded a single positive cluster with a significant processing difference between violated and matching sound-pseudowords ($p = 0.014$). This difference was found between 406 – 512 ms after the onset of the pseudoword with a broad topological distribution during peak cluster activation. No clusters of significant differences were found for the later TW (see Figure 4A).

For the athlete group, no clusters of significant processing differences were found between the violated and matching sound-pseudoword pairs in the testing phase for either the N400 or the later TW (see Figure 4B).

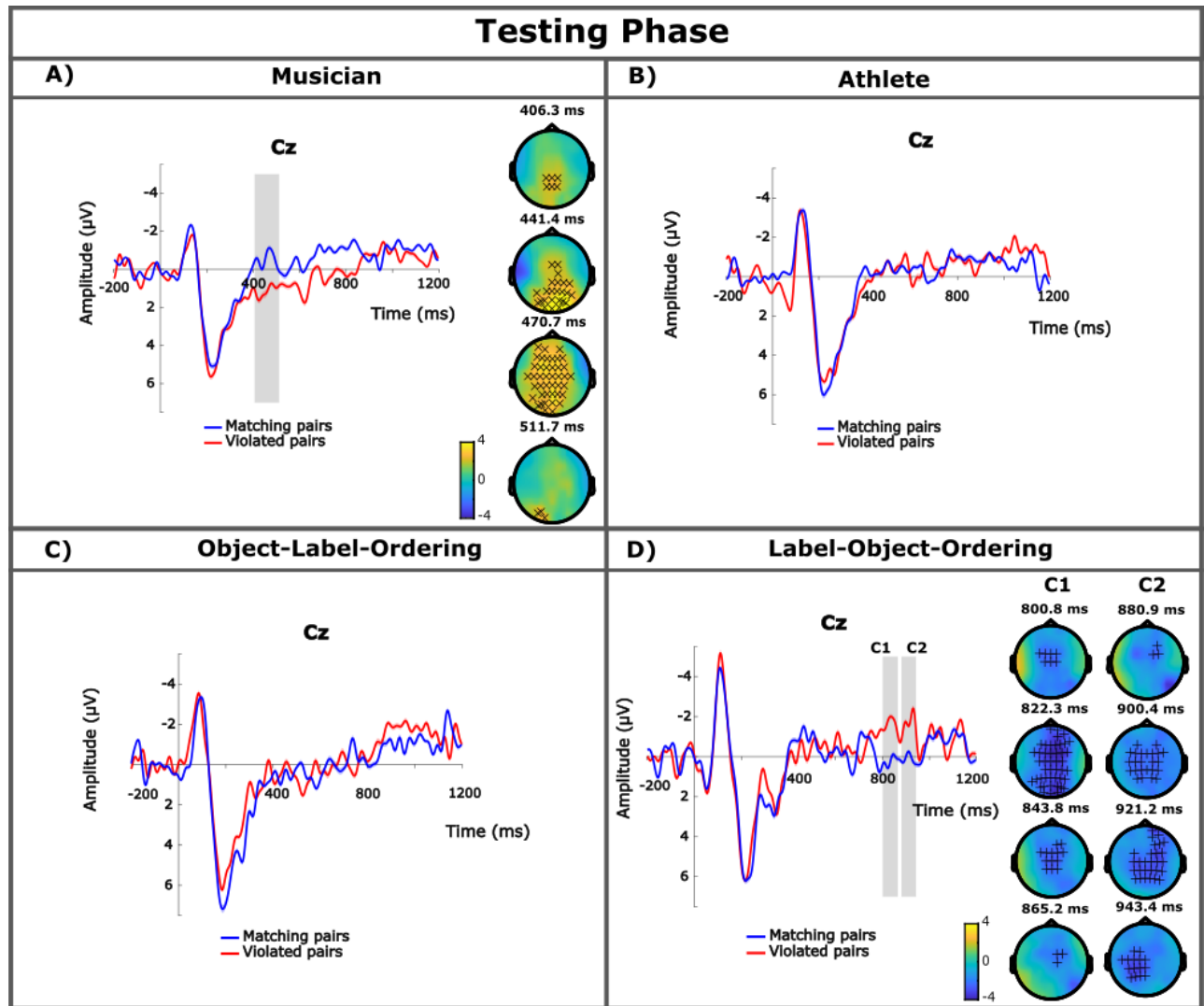


Figure 4: Testing phase results. ERPs are presented time-locked to the onset of the second stimuli in the pairing for the testing phase of each experiment (A, B, C represent pseudowords, D represents environmental sounds). The Cz electrode was selected as a representative electrode for all of the experiments. The blue lines indicate the matching pairs while the red lines indicate the violated pairs. The lighter-shaded lines indicate the standard error of the waveform. Gray bars are used to illustrate the time window where significant differences were found between the conditions. In the case of multiple clusters, the clusters are labeled C1 and C2. Topological difference maps from the cluster-based permutation test are presented only where significant differences (or trends) were found and show the electrode distribution in the cluster over time (the + symbol indicates $0.1 > p > 0.5$ and the X symbol indicates $p < 0.05$). Results of Object-Label-Ordering (C) are reprinted from a previous publication (Casper et al., 2022).

3 Experiment 2: object-label-ordering and label-object-ordering

3.1 Materials and methods

3.1.1 Participants

Participants in the object-label-ordering group were 22 students from a previous experiment (auditory sequential; Cospers et al., 2022). All participants (16 female) were included in the final dataset, aged 19 to 32 years of age ($M = 23.86$ years; $SD = 3.4$ years). Table 1 provides an overview of all participants.

For the label-object-ordering group, 24 students were recruited to participate in the study. One participant ($n = 1$) was removed for not meeting inclusion criteria, leaving 23 students (11 female) in the final dataset, aged 18 to 29 years of age ($M = 23.36$ years; $SD = 5.23$ years). For an overview of all participants, see Table 1. None of the participants of this experiment had participated in any other auditory associative word learning experiment. The recruitment of participants as well as all inclusion and exclusion criteria were identical to experiment 1.

3.1.2 Stimuli and Procedure

The procedure of experiment 2 followed the procedure of experiment 1. Note, the example stimuli combinations in Figure 1A and 1B are provided for experiment 1 as well as for the object-label-ordering group of this experiment. The pairings for the label-object-ordering group were presented in the opposite order. For the object-label-ordering group, environmental sounds were first presented, followed by an inter-stimulus pause of 600 ms, then a pseudoword was presented (identical to experiment one; see Figure 1C). For the label-object-ordering group, pseudowords were presented first, then a 600 ms inter-stimulus pause, followed by the environmental sound (see Figure 1D). The data for the object-label-ordering group is from a previous publication (auditory sequential; Cospers et al., 2022) reporting the same findings for means of comparison; however, the data have been newly analyzed for the later time window (800 – 1200 ms) in both the training and testing phases to be comparable to the other groups in experiment 1 and 2.

3.2 Results

3.2.1 Behavioral Results

The object-label-ordering group of the previous study did not indicate any behavioral results above chance (auditory sequential; Cospers et al., 2022). For the label-object-ordering group, a

one-sample t-test indicated that the participants were significantly above chance in identifying matching and violated pseudoword-sound pairings in the testing phase at 73% ($M = 23.36$ correct responses; $SD = 5.03$ responses), $t(22) = 6.7677$, $p < 0.001$ (see Figure 2).

3.2.2 EEG: Training Phase

The previous study indicated a single cluster of significant differences in the second half of the training phase for the object-label-ordering group (auditory sequential; Cosper et al., 2022). No clusters of significant processing differences were found in the later TW in either the first or second half of training (see Figure 3C).

The analysis of both the first and second half of training for the label-object-ordering group yielded no clusters in which significant differences between inconsistent and consistent pseudoword-sound pairings were found in either TW (see Figure 3D).

3.2.3 EEG: Testing Phase

As previously reported, the object-label-ordering group, no clusters in which significant processing differences between matching and violated sound-pseudoword pairs were found in either the N400 TW (auditory sequential; Cosper et al., 2022). Further analysis similarly found no clusters of significant processing differences in the later TW (see Figure 4C).

The analysis of the label-object-ordering group did not yield any clusters in which significant differences were found between violated and matching pseudoword-sound pairs in either the 400 – 800 ms or the 800 – 1200 ms TW. However, two trends were identified in the later TW. Cluster 1, from 801 – 865 ms after the onset of the sound, narrowly escaped significance differences between processing of violated and matching pairings ($p = 0.068$). Cluster 2, from 881 – 943 ms, also yielded a trend in the differences between the processing of violated over matching pairs ($p = 0.074$). These clusters in which a trend has been identified between the conditions in the testing phase can be seen in Figure 4D.

4 Discussion

To improve our understanding of external and internal modulators of associative auditory word learning in adults we conducted two behavioral and ERP experiments in which novel word forms were sequentially linked to novel environmental sounds. First, two expert groups were

tested on object-label ordered learning sequences. Second, a non-expert group was tested on label-object ordered sequences and compared to a non-expert group previously tested on object-label ordered sequences (Casper et al., 2022). For musicians, electrophysiological analysis of the training phase did not reveal an N400 effect, and at test, behavioral response accuracy was not above chance. However, we observed a late EEG effect in the testing phase with more positive-going ERP responses for violated than matching stimulus pairs. In the athlete group, neither behavioral nor ERP analyses of the training and testing phases indicated successful auditory associative word learning. For the two non-expert groups, label-object-ordering revealed an above-chance response accuracy in the testing phase as well as a marginally significant late negativity for violated than matching pairs in the testing phase. In contrast object-label-ordering did not indicate above-chance response accuracy or yield any ERP effects during the testing phase (cf. Casper et al., 2022). These results thus suggest that both experimentally manipulated learning factors, auditory expertise and auditory saliency from stimulus ordering, impacted on the learnability of stimulus pairs. We will discuss the potential impact of these factors on the behavioral and electrophysiological outcome of auditory associative learning in more detail below.

For the testing phase, musicians showed a positive cluster of electrodes for the processing of violated over matching sound-pseudoword pairs. Although the latency of this effect generally aligns with the expectation of an N400 (Bechtold et al., 2023; Casper et al., 2022; Kutas & Federmeier, 2000, 2011), given absolute latency and time window limitations of cluster-based permutation analyses (Sassenhagen & Draschkow, 2019), the polarity is reversed and does not mirror the polarity for the elicited component to be an N400. We suggest that the elicited effect is a reflection of the late positive component (LPC), which is typically found in musicians as a response to incongruous elements in musical melody and rhythm, increasing with familiarity and musical expertise (cf. Besson & Faïta, 1995; Calma-Roddin & Drury, 2020; Featherstone et al., 2013; Miranda & Ullman, 2007). In their analysis, Featherstone and colleagues (2013) reported an LPC (also related to the P600 component for syntactic processing) in musicians between 500 – 700 ms after the onset of the target stimulus. The authors, moreover, reported that non-musicians exhibited an N500 response (related to the N400 for semantic processing) for the same stimuli presentations. Similarly, Besson and Faïta (1995) reported an LPC (400 – 800 ms) for congruity with larger amplitude for musicians than for non-musicians. The interpretation of these findings is that individuals with musical expertise (i.e., musicians) process incongruent semantic relations in musical stimuli differently than non-musicians and that the neural underpinnings of these processes may change with expertise (Besson, 1998;

Besson et al., 1994; Featherstone et al., 2013; Miranda & Ullman, 2007; Steinbeis et al., 2006). Thus, in our study, it may not only be the case that musicians implicitly learned a link between auditory objects and labels, but musicians may also have processed these associations differently than non-musicians. As no behavioral learning effects were found, we interpret the LPC effect as an indication of implicit learning that does not necessarily need to be lexical-semantic in nature. As N400 effects have been found with musicians in associative word learning with visual objects (e.g., Dittinger et al., 2016, 2017, 2019), it is unlikely that the LPC simply reflects word learning in musicians. Instead, the LPC may simply reflect a process akin to auditory statistical learning. This suggestion is supported by experiments in which two auditory elements were sequentially linked, and in which similar positivities were found (e.g., Friederici et al., 2011; Mandikal Vasuki et al., 2017). Thus, the present results are consistent with previous findings in musicians evidencing enhanced auditory processing and memory (Chan et al., 1998; Cohen et al., 2011; George & Coch, 2011; Ho et al., 2003; Paraskevopoulos et al., 2017), and the transfer of musical expertise to language processing (Dittinger et al., 2016, 2017, 2019; Pesnot Lerousseau & Schön, 2021; Schön & Francois, 2011).

In contrast to the musicians, no significant effects were found in the athlete group in either behavioral or ERP analyses. Athletes were not above chance levels at identifying correct and incorrect sound-pseudoword pairings. Furthermore, electrophysiological brain signals did not indicate any significant processing differences between matching and violated sound-pseudoword pairings during the testing phase. On the one hand, these findings corroborate and replicate the findings of Cospers and colleagues (2022) showing that adults (without auditory expertise) are not able to map spoken labels onto auditory environmental sounds in sequential object-label ordered presentation. On the other hand, the absence of findings for the athlete group also support the findings of Heppe and colleagues (2016) as well as Zoudji and Thon (2003) in that the attentional and cognitive advantages of athletes do not transfer beyond sport-related topics.

An important outcome of the present study was the non-replication of the N400 effect in the training phase that had been reported in an adult study of word learning by Cospers and colleagues (2022). In this previous study, the N400 effect in the training phase was interpreted as being related to the sequentiality of the stimulus presentation, whereas sequential presentation yielded an N400 response in training and simultaneous presentation did not yield an N400 response. The present results from the three ERP experiments including the object-label-order condition (auditory sequential; Cospers et al., 2022) suggest that our previous interpretation might have been inadequate. The non-replication of the N400 effect could be

either the result of insufficient power (yet participant numbers were comparable) or hinting to the possibility that there are other factors influencing the presence of the N400 effect during associative word learning which is expected to occur only in the second half of the training phase. Such factors could be a high sensitivity to the dynamics of learning (for example, if sounds and words are linked only in the very last repetitions, the effect would be not observable) or subtle differences in motivation and attention on the side of the participants. In order to dissect the influences behind the N400 during learning, further experiments are needed and should be the focus of future research.

Regarding stimulus order, the behavioral analysis of the testing phase yielded above-chance accuracies only for the label-object ordered but not for the reverse condition. This suggests that adults are better learners in in auditory-sequential associative word learning, when the more salient auditory label appears first (cf. Cospers et al., 2022). Despite above-chance behavioral responses, the ERP analysis of the testing phase for the label-object-ordering group did not meet the threshold of significance for processing differences between matching over violated pseudoword-sound pairings. Two clusters were identified with nearing statistically significant processing differences as identified by an N400-like component. Although these trends appear much later than the target TW of the N400, later N400-like responses have been seen in both children (for a review, see Junge et al., 2021) and in adults (auditory-simultaneous associative word learning; Cospers et al., 2022). Alternatively, the identification of the sounds could have taken longer given that the sounds were not as initially unique in their identification as the pseudowords. Even if sound identification did not always take longer, but might have been more variable, this could have led to the increased latency and non-significance of the effect which may have captured just the final processing stage in which all the unique sounds were identified. Note that it is unlikely that the environmental sounds per se yielded a later effect. Orgs and colleagues (2006) provided evidence that contextual priming between written words and environmental sounds elicited an N400 effect with an even earlier onset than a picture and a written word. However, the environmental sound stimuli used by Orgs and colleagues (2006) were much shorter than the environmental sound stimuli used in the current set of experiments (300 ms versus 950 ms, respectively). It is therefore possible that the associative memory processes, be it from the encoding process of association between the two stimuli presented or from the maintenance of the environmental sound, may have contributed to the delay of the N400 in the object-label-ordering condition. Thus, the late negative component in the label-object-ordering group might similarly reflect similar a violation of lexico-semantic expectation,

as suggested in Cospers and colleagues (2022), but this proposal would need to be confirmed in an independent study.

In sum, given the advantages in behavioral response accuracy and the observed trends in the ERP analysis of the testing phase, we suggest that auditory-sequential associative word learning does benefit from label-object-ordering in comparison to object-label-ordering in a general, non-musician population. This may be due to the increased saliency and memorizability of words compared to sounds which may result in a better matching of words and sounds when the sounds are presented in the second position in the sequence (cf. Renvall et al., 2021). In contrast, when sounds are presented first, it may be harder to form a memory representation and thus, when the word appears in the second position, matching may be difficult because the sound has already faded.

We caution to the overgeneralization of our results, as we are limited in our scope of interpretation. Although musicians and athletes were tested as expertise groups, groups assignment was only substantiated by the self-reports of the individual participants and no measures were conducted to identify the expertise of either group. In future studies, this should be addressed, for example, by investigating individual indexes of modality-specific (i.e., visual and auditory) short-term memory recognition accuracies with both verbal and nonverbal stimuli. Additionally, not all of the musicians were professional musicians. Similarly, athletes were categorized as recreational-athletes and no semi-professional, professional, or elite athletes were identified in the study. Distinguishing between multiple within-group levels may yield further findings. Furthermore, both expertise groups only participated in learning from object-label-ordering. Given the overall advantage of musicians in auditory processing (Chan et al., 1998; Cohen et al., 2011; George & Coch, 2011; Ho et al., 2003; Paraskevopoulos et al., 2017) and the increased salience and accuracy of short-term recognition memory of verbal over nonverbal stimuli (Calignano et al., 2021; Lukics & Lukács, 2022; Renvall et al., 2021; Talamini et al., 2022), musicians may have provided further evidence of general advantages of salient stimulus ordering in auditory-sequential associative word learning.

Finally, we will discuss the relative difficulty of associative word learning in the auditory domain as compared to the visual domain. In Cospers and colleagues (2022), visual behavioral performance accuracy was over 80%, while the auditory experiments yielded either no learning success (auditory sequential) or success rates below 70% (auditory simultaneous 66%). This seems to indicate that learning words for sounds in an associative manner is more difficult compared to learning words for visual objects. Beside the advantages of multimodal learning

over unimodal learning and the integration of modalities in a multimodal language system (e.g., Eördegh et al., 2019; Matusz et al., 2015; Vigliocco et al., 2014), one potential further explanation for the visual advantage in relation to the specific type of stimuli used in our studies might be that the two stimuli clearly originate from different sound sources: environmental sounds vs. human speech. Auditory sequential processing does not only serve to discover links between all stimuli that naturally occur, but also to determine the sound sources of acoustic streams in so-called auditory scene analysis (Bregman, 1990; Winkler & Schröger, 2015). This might potentially be a more general problem in unimodal learning contexts, namely that the requirement to link two sounds competes with the necessity to segregate two sounds and assign them to different sources. Together with the general human preference for visual stimuli in adulthood, this might contribute to the relative difficulty of learning words for sounds, also reflected in the lower frequency of sound-related words compared to visual words (cf. Chedid et al., 2019; Chen et al., 2019; Lynott et al., 2020; Majid et al., 2018; Miklashevsky, 2018; Speed & Brybaert, 2022; Vergallito et al., 2020).

5 Conclusion

In the current study, we have shown that adult musicians do not have a behavioral advantage over non-musicians and athletes in auditory-sequential associative word learning. However, musicians do display evidence of implicit learning, albeit the learning may not be the result of lexico-semantic relationships. Furthermore, there is a behavioral advantage and the indication of a possible electrophysiological effect for label-object-ordering over object-label-ordering in auditory-sequential associative word learning. Our findings show how musical expertise and the salience of auditory-verbal stimuli influence association of sounds and labels. Future investigations into modality-specific associative learning should consider individual expertise as a factor in learning as well as saliency and memorizability of the stimuli in a given modality.

Funding

Data was collected at the University of Osnabrück, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number GRK-2185/1 (DFG Research Training Group Situated Cognition)

Data Availability Statement:

The data and scripts used for the analysis can be found here: <https://osf.io/8wk9d/> and the experiments were not preregistered.

Author contributions:

Samuel H. Cospers: Conceptualization, Formal analysis, Investigation, Data Curation, Writing – Original draft, Visualization. **Claudia Männel:** Conceptualization, Writing – Review & editing, Supervision. **Jutta L. Mueller:** Conceptualization, Writing – Review & editing, Supervision.

Acknowledgements:

The authors would like to thank N. Gockel, L. Steinbach, C. Dammeyer, and all students and student research assistants of the Kindersprachlabor for their help in experiment coding and data collection. We are especially thankful to all of the participants.

6 Bibliography

- Alves, H., Voss, M., Boot, W., Deslandes, A., Cossich, V., Inacio Salles, J., & Kramer, A. (2013). Perceptual-Cognitive Expertise in Elite Volleyball Players. *Frontiers in Psychology, 4*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00036>
- Ballas, J. A., & Howard, J. H. (1987). Interpreting the Language of Environmental Sounds. *Environment and Behavior, 19*(1), 91–114. <https://doi.org/10.1177/0013916587191005>
- Bechtold, L., Cospers, S. H., Malyshevskaya, A., Montefinese, M., Morucci, P., Nicolai, V., Repetto, C., Zappa, A., & Shtyrov, Y. (2023). Brain Signatures of Embodied Semantics and Language: A Consensus Paper. *Journal of Cognition, 6*(1), 61. <https://doi.org/10.5334/joc.237>
- Besson, M. (1998). Meaning, structure, and time in language and music. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 17*, 921–950.
- Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1278–1296. <https://doi.org/10.1037/0096-1523.21.6.1278>
- Besson, M., Faïta, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters, 168*(1), 101–105. [https://doi.org/10.1016/0304-3940\(94\)90426-X](https://doi.org/10.1016/0304-3940(94)90426-X)
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press. <https://doi.org/10.7551/mitpress/1486.001.0001>
- Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage, 25*(3), 958–968. <https://doi.org/10.1016/J.NEUROIMAGE.2004.12.019>
- Calignano, G., Valenza, E., Vespignani, F., Russo, S., & Sulpizio, S. (2021). The unique role of novel linguistic labels on the disengagement of visual attention. *Quarterly Journal of Experimental Psychology, 74*(10), 1755–1772. <https://doi.org/10.1177/17470218211014147>
- Calma-Roddin, N., & Drury, J. E. (2020). Music, Language, and The N400: ERP Interference Patterns Across Cognitive Domains. *Scientific Reports, 10*(1), Article 1. <https://doi.org/10.1038/s41598-020-66732-0>

- Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature*, 396(6707), Article 6707. <https://doi.org/10.1038/24075>
- Chedid, G., Brambati, S. M., Bedetti, C., Rey, A. E., Wilson, M. A., & Vallet, G. T. (2019). Visual and auditory perceptual strength norms for 3,596 French nouns and their relationship with other psycholinguistic variables. *Behavior Research Methods*, 1–12. <https://doi.org/10.3758/s13428-019-01254-w>
- Chen, I.-H., Zhao, Q., Long, Y., Lu, Q., & Huang, C.-R. (2019). Mandarin Chinese modality exclusivity norms. *PLOS ONE*, 14(2), e0211336. <https://doi.org/10.1371/journal.pone.0211336>
- Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic Bulletin and Review*, 18(3), 586–591. <https://doi.org/10.3758/s13423-011-0074-0>
- Cosper, S. H., Männel, C., & Mueller, J. L. (2020). In the absence of visual input: Electrophysiological evidence of infants' mapping of labels onto auditory objects. *Developmental Cognitive Neuroscience*, 45, 100821. <https://doi.org/10.1016/j.dcn.2020.100821>
- Cosper, S. H., Männel, C., & Mueller, J. L. (2022). Mechanisms of associative word learning: Benefits from the visual modality and synchrony of labeled objects. *Cortex*, 152, 36–52. <https://doi.org/10.1016/J.CORTEX.2022.03.020>
- 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>
- Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations. *Journal of Cognitive Neuroscience*, 28(10), 1584–1602. https://doi.org/10.1162/jocn_a_00997
- Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast Brain Plasticity during Word Learning in Musically-Trained Children. *Frontiers in Human Neuroscience*, 11(May), 233. <https://doi.org/10.3389/fnhum.2017.00233>
- Dittinger, E., Scherer, J., Jäncke, L., Besson, M., & Elmer, S. (2019). Testing the influence of musical expertise on novel word learning across the lifespan using a cross-sectional approach in children, young adults and older adults. *Brain and Language*, 198, 104678. <https://doi.org/10.1016/j.bandl.2019.104678>

- Eördegh, G., Óze, A., Bodosi, B., Puszta, A., Pertich, Á., Rosu, A., Godó, G., & Nagy, A. (2019). Multisensory guided associative learning in healthy humans. *PLOS ONE*, *14*(3), e0213094. <https://doi.org/10.1371/journal.pone.0213094>
- Featherstone, C. R., Morrison, C. M., Waterman, M. G., & MacGregor, L. J. (2013). Semantics, Syntax or Neither? A Case for Resolution in the Interpretation of N500 and P600 Responses to Harmonic Incongruities. *PLOS ONE*, *8*(11), e76600. <https://doi.org/10.1371/journal.pone.0076600>
- Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to Natural Grammar Learning: Preliminary Evidence from 4-Month-Old Infants. *PLOS ONE*, *6*(3), e17920. <https://doi.org/10.1371/journal.pone.0017920>
- 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
- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094. <https://doi.org/10.1016/j.neuropsychologia.2011.02.001>
- Heppe, H., Kohler, A., Fleddermann, M.-T., & Zentgraf, K. (2016). The Relationship between Expertise in Sports, Visuospatial, and Basic Cognitive Skills. *Frontiers in Psychology*, *7*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00904>
- Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*, 439–450. <https://doi.org/10.1037/0894-4105.17.3.439>
- 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>
- 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>
- Junge, C., Boumeester, M., Mills, D. L., Paul, M., & Cosper, S. H. (2021). Development of the N400 for Word Learning in the First 2 Years of Life: A Systematic Review. *Frontiers in Psychology*, *12*, 689534. <https://doi.org/10.3389/fpsyg.2021.689534>

- Klein, D. J., König, P., & Körding, K. P. (2003). Sparse Spectrotemporal Coding of Sounds. *EURASIP Journal on Advances in Signal Processing*, 2003(7), 902061. <https://doi.org/10.1155/S1110865703303051>
- 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>
- Lukics, K. S., & Lukács, Á. (2022). Modality, presentation, domain and training effects in statistical learning. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-24951-7>
- Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, 52(3), 1271–1291. <https://doi.org/10.3758/s13428-019-01316-z>
- Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O’Grady, L., Woll, B., LeLan, B., de Sousa, H., Cansler, B. L., Shayan, S., de Vos, C., Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemans, M., Ozturk, O., ... 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>
- Mandikal Vasuki, P. R., Sharma, M., Ibrahim, R., & Arciuli, J. (2017). Statistical learning and auditory processing in children with music training: An ERP study. *Clinical Neurophysiology*, 128(7), 1270–1281. <https://doi.org/10.1016/j.clinph.2017.04.010>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/J.JNEUMETH.2007.03.024>
- Matusz, P. J., Thelen, A., Amrein, S., Geiser, E., Anken, J., & Murray, M. M. (2015). The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *European Journal of Neuroscience*, 41(5), 699–708. <https://doi.org/10.1111/ejn.12804>
- 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>

- Miklashevsky, A. (2018). Perceptual experience norms for 506 Russian nouns: Modality rating, spatial localization, manipulability, imageability and other variables. *Journal of Psycholinguistic Research*, 47(3), 641–661. <https://doi.org/10.1007/s10936-017-9548-1>
- 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>
- Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage*, 38(2), 331–345. <https://doi.org/10.1016/j.neuroimage.2007.07.034>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869. <https://doi.org/10.1155/2011/156869>
- Orgs, G., Lange, K., Dombrowski, J.-H., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, 62(3), 267–272. <https://doi.org/10.1016/j.bandc.2006.05.003>
- Paraskevopoulos, E., Chalas, N., & Bamidis, P. (2017). Functional connectivity of the cortical network supporting statistical learning in musicians and non-musicians: An MEG study. *Scientific Reports*, 7(1), Article 1. <https://doi.org/10.1038/s41598-017-16592-y>
- Pesnot Lerousseau, J., & Schön, D. (2021). Musical Expertise Is Associated with Improved Neural Statistical Learning in the Auditory Domain. *Cerebral Cortex*, 31(11), 4877–4890. <https://doi.org/10.1093/cercor/bhab128>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.r-project.org/>
- 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
- Renvall, H., Seol, J., Tuominen, R., Sorger, B., Riecke, L., & Salmelin, R. (2021). Selective auditory attention within naturalistic scenes modulates reactivity to speech sounds. *European Journal of Neuroscience*, 54(10), 7626–7641. <https://doi.org/10.1111/ejn.15504>
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), e13335. <https://doi.org/10.1111/psyp.13335>

- 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>
- Schön, D., & Francois, C. (2011). Musical Expertise and Statistical Learning of Musical and Linguistic Structures. *Frontiers in Psychology*, 2. <https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00167>
- 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.COGLING.2017.06.001>
- Speed, L. J., & Brybaert, M. (2022). Dutch sensory modality norms. *Behavior Research Methods*, 54(3), 1306–1318. <https://doi.org/10.3758/s13428-021-01656-9>
- Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses. *Journal of Cognitive Neuroscience*, 18(8), 1380–1393. <https://doi.org/10.1162/jocn.2006.18.8.1380>
- Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2022). Auditory and visual short-term memory: Influence of material type, contour, and musical expertise. *Psychological Research*, 86(2), 421–442. <https://doi.org/10.1007/s00426-021-01519-0>
- 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>
- Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, 34(6), 1093–1106. <https://doi.org/10.1111/j.1551-6709.2010.01118.x>
- van Petten, C., & Rieffers, 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)
- Vanek, N., Sóskuthy, M., & Majid, A. (2021). Consistent verbal labels promote odor category learning. *Cognition*, 206, 104485. <https://doi.org/10.1016/j.cognition.2020.104485>
- Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods*, 1–18. <https://doi.org/10.3758/s13428-019-01337-8>

- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130292. <https://doi.org/10.1098/rstb.2013.0292>
- Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes ‘expert’ in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology*, 24(6), 812–826. <https://doi.org/10.1002/acp.1588>
- 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>
- Winkler, I., & Schröger, E. (2015). Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain and Language*, 148, 1–22. <https://doi.org/10.1016/j.bandl.2015.05.003>
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272. <https://doi.org/10.1037/0033-295X.114.2.245>
- Yan, Y., Yang, Y., Ando, M., Liu, X., & Kambara, T. (2021). Multisensory Connections of Novel Linguistic Stimuli in Japanese as a Native Language and Referential Tastes. *European Journal of Investigation in Health, Psychology and Education*, 11(3), Article 3. <https://doi.org/10.3390/ejihpe11030074>
- Zoudji, B., & Thon, B. (2003). Expertise and Implicit Memory: Differential Repetition Priming Effects on Decision Making in Experienced and Inexperienced Soccer Players. *International Journal of Sport Psychology*, 34, 189–207.