Neuroscience Research 64 (2009) 185–190

Contents lists available at ScienceDirect

Neuroscience Research

journal homepage: www.elsevier.com/locate/neures

Visual statistical learning of shape sequences: An ERP study

Dilshat Abla [a,b], Kazuo Okanoya [a,*]

[a] Laboratory for Biolinguistics, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[b] Xinjiang Medical University, Urumqi, Xinjiang, PR China

ARTICLE INFO

Article history:
Received 15 October 2008
Received in revised form 13 February 2009
Accepted 25 February 2009
Available online 13 March 2009

Keywords:
Statistical learning
Word segmentation
Visual shape
Event-related potentials
N400

ABSTRACT

Behavioral experiments have found that infants and adults learn statistically defined patterns presented in auditory and visual input sequences in the same manner regardless of whether the input was linguistic (syllables) or nonlinguistic (tones and shapes). In order do determine the time course and neural processes involved in online word segmentation and statistical learning of visual sequence, we recorded event-related potentials (ERPs) while participants were exposed to continuous sequences with elements organized into shape-words randomly connected to each other. After viewing three 6.6 min sessions of sequences, the participants performed a behavioral choice test. The participants were divided into two groups (high and low learners) based on their behavioral performance. The overall mean performance was 72.2%, indicating that the shape sequence was segmented and that the participants learned the shape-triplets statistically. Grand-averaged ERPs showed that triplet-onset (the initial shapes of shape-words) elicited larger N400 amplitudes than did middle and final shapes embedded in continuous streams during the early learning sessions of high learners, but no triplet-onset effect was found among low learners. The results suggested that the N400 effect indicated online segmentation of the visual sequence and the degree of statistical learning. Our results also imply that statistical learning represents a common learning device.

1. Introduction

Behavioral experiments have reported that humans readily learn statistically defined patterns in auditory sequences (Goodsitt et al., 1993; Christophe et al., 1994; Jusczyk and Aslin, 1995; Saffran et al., 1996a,b, 1999; Aslin et al., 1998; Christiansen et al., 1998; Brent, 1999; Pena et al., 2002; Seidenberg et al., 2002) and visual input sequences (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005, 2008). Saffran et al. (1996a,b, 1999) have suggested that both infants and adults can detect and use the statistical properties of linguistic (syllable) and non-linguistic (tone) sequences to discriminate among "words" presented in continuous sequences. Some have hypothesized that statistical information is a particularly useful cue in regard to word segmentation (Saffran et al., 1996a, 1999; Aslin et al., 1998). Kirkham et al. (2002) investigated the domain generality of statistical learning in infancy by examining whether young infants were able to extract statistical information from visual stimuli. They habituated infants to sequences of discrete visual stimuli ordered according to a statistically predictable pattern. The infants subsequently viewed the familiar pattern alternating with a novel sequence of identical stimulus components. The infants exhibited significantly greater interest in the novel sequence at all ages, suggesting that they were able to extract the statistical information in the visual shape sequences that defined word boundaries in the same way in which they processed auditory stimuli. Although these behavioral studies can address outcome of statistical learning by using exposure-test design, they do not provide information about the neural processes that subserve the learning process itself.

In a recent ERP study using an online tone-word segmentation task, Abla et al. (2008) found an ERP component that reflected tone-word segmentation and process of statistical learning. They recorded ERPs while adults listened to a continuous auditory stream in which six tritone nonsense words were embedded according to a statistical rule (Saffran et al., 1999). The most likely cues for the beginnings and ends of the tone-words were the transitional probabilities (TPs) between tones. The TPs were lower between than within the tritone words. After listening to three 6.6 min sessions of sequences, their participants performed a behavioral choice test, in which they were instructed to indicate the most familiar tone sequence in each test trial by pressing buttons. The overall mean performance was above 70%, indicating that the tone sequence was segmented and that the participants learned the tone words statistically. Their ERP results showed that triplet-word onset (the initial tones of tone-words, the less predictable position) elicited larger N400 amplitudes than did middle and final tones embedded in continuous streams, and the

* Corresponding author. Tel.: +81 48 467 7502; fax: +81 48 467 7503.
E-mail address: okanoya@brain.riken.jp (K. Okanoya).

0168-0102/$ – see front matter © 2009 Elsevier Ireland Ltd and the Japan Neuroscience Society. All rights reserved.
doi:10.1016/j.neures.2009.02.013

amplitude of the N400 was highly correlated with the TPs, with the lowest tone stream TP positions eliciting larger potentials. The N400 effect was larger in their higher learner groups and significantly differed between the three learning sessions, but no word-onset effect was found among low learners. Sanders et al. (2002) and Cunillera et al. (2006) conducted similar studies. These researchers recorded ERPs while adults listened to several trisyllabic, pronounceable nonsense-words (*babupu*, *bupada*, *dutaba*. . .) presented as continuous speech as per the experimental procedure followed by Saffran et al. (1996a). They also found that the onset of nonsense words elicited larger ERP components than did the middle or final syllables. These studies clearly demonstrated that the N400 component is elicited during the learning of nonsense auditory sequences and can be regarded as indicators of statistical segmentation.

The N400 was described for the first time by Kutas and Hillyard (1980) in response to incongruent sentence endings, such as 'He spreads the warm bread with socks.' N400 was proposed to reflect the interruption of ongoing sentence processing by a semantically inappropriate word. This early hypothesis was soon abandoned when it was shown that large N400s could also be evoked by congruent but unexpected sentence endings, such as 'He was soothed by the gentle wind.' The more expected ending elicited the smaller amplitude of the N400 (Kutas and Hillyard, 1984). Later, N400 effects were also found with semantic priming in lexical tasks involving isolated word pairs (Bentin et al., 1985), with meaningful stimuli other than words, such as faces (Barrett and Rugg, 1989), notably objects (Holcomb and McPherson, 1994) and environmental sounds (Van Petten and Rheinfelder, 1995). There are likely to be multiple latent components associated with a variety of cognitive processes that contribute to the N400 observed at the scalp. Perceptual segmentation is probably one of several underlying cognitive factors contributing to the observed N400.

Regarding the word segmentation, previous ERP studies have used only auditory streams as stimuli, thus providing no data on visual stimuli. Whether the N400 component indexing word segmentation is elicited during visual shape segmentation as well as during an auditory segmentation task remains unknown. The ERP indicator reflecting word segmentation might also be observed during segmentation of visual shape sequences, since use of the statistical consistencies among adjacent elements to create perceptual groups should represent a common learning device.

The current study used ERPs to elucidate the neural correlates of the online segmentation and statistical learning of visual shape sequences. We recorded 32-channel visual ERPs while adult participants were exposed to continuous nonlinguistic visual sequences consisting of elements organized into "shape-triplets" in three 6.6 min sessions, which was the same experimental design with the previous auditory study of Abla et al. (2008). Using the sequence stimuli from the three sessions, we were able to compare the ERPs elicited by the continuous stimuli when they were and were not segmented as triplets, thereby elucidating the neural processes corresponding to online statistical learning of visual stimuli. By using nonsense shape triplets, we also sought to identify basic perceptual processes that operate independently from linguistic faculty.

2. Materials and methods

2.1. Participants

Eighteen adults (8 males, 10 females) with normal vision participated in this experiment. The ages of participants ranged from 20 to 39 years ($M = 23.7, \pm 5.1$). All participants were right-handed (as per the Edinburgh Inventory, Oldfield, 1971) with no history of neurological disease. The RIKEN Ethics Committee approved all procedures in advance and all participants provided written informed consent before each experiment.

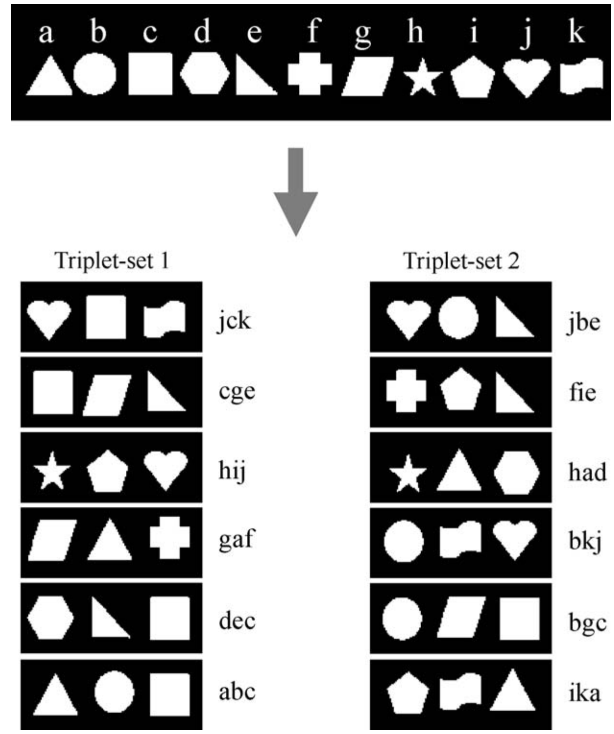

Fig. 1. The 11 shapes and shape-words used in the present experiment.

2.2. Stimuli

We generated 11 white shapes (triangle, circle, square, hexagon, right-triangle, cross, diamond, pentangle, trapezoid, heart-shape, and flag-shape, identified as a, b, c, d, e, f, g, h, i, j, and k, respectively; see Fig. 1) on a black background. The maximum height and width of the shapes were scaled to be equal. Three shapes were combined to form one triplet, yielding six shape-triplets in each of two triplet-sets (Triplet-set 1, T1: jck, cge, hij, gaf, dec, abc; Triplet-set 2, T2: jbe, fie, had, bkj, bgc, ika) (Fig. 1). The statistical structure of these triplets exactly mirrored that of the tone-words used by Saffran et al. (1999) and Abla et al. (2008). For T1, the TPs between shapes within triplets averaged 0.64 (range = 0.25–1.00), while the TPs between triplets averaged 0.14 (range = 0.05–0.60). The statistical structure of T2 was very similar to that of T1. TPs between shapes within triplets averaged 0.71 (range = 0.33–1.00) while TPs between triplets averaged 0.18 (range = 0.07–0.53).

The six triplets were presented in random order, with no temporal gaps between triplets, to produce a 6.6 min continuous stream of shapes (e.g., *jckhijcgejckgafhijabcdec*. . .). Each triplet was repeated 40 times in one stream, and a particular triplet was never immediately repeated. We constructed the shape sequence using Gentask of Stim software (Neuro Scan, Inc.).

We developed 36 pair-items and followed the test used by Saffran et al. (1999) to assess learning. Each test item consisted of two tri-shape sequences: one familiar triplet and one new triplet (as nonword in Saffran et al., 1999). If participants observed T1 in the continuous presentation sessions, the familiar triplets were extracted from T1 and the new triplets were extracted from T2, and vice versa. All six triplets from each triplet-set were paired with those from the other triplets, generating 36 test trials. The two stimuli (familiar triplet and new triplet), separated by a 1.2 s pause, were randomly presented in each trial. We used an intertrial interval of 3–4 s.

2.3. Procedure

We asked the participants to view the continuous shape stimuli on a monitor placed 120 cm in front of them without moving their eyes. We also asked them to relax and to avoid consciously analyzing the shape sequences as they viewed them. The participants were not given any information regarding the nature of the stimuli, including the statistical structure of the shape sequences, and were not told which triplet-set they would observe.

Each shape was presented for 550 ms, as the previous study of tone presentation (Abla et al., 2008), in a continuous stream, with no break or delay between shapes; stimuli measured from 4 to 5 cm in height and width (1.9–2.4°). Participants then viewed the 6.6 min sequence of one shape-stream (T1 or T2) described above followed by a short break. The stimulus presentation was repeated three times. Each of the three 6.6 min streams was randomized separately. The participants completed the behavioral choice test after a total of 19.8 min viewing time. We