Title
Predictive Arm Placement in the Statistical Learning of Position Sequences

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN
1069-7977

Authors
Dale, Rick
Duran, Nicholas

Publication Date
2009

Peer reviewed

Predictive Arm Placement in the Statistical Learning of Position Sequences

Nicholas D. Duran (nduran@memphis.edu) **Rick Dale (radale@memphis.edu)**
Department of Psychology, The University of Memphis, Memphis, TN, 38152

Abstract

Researchers have shown that statistical learning is a pervasive and deeply entrenched human ability. As the input environment becomes more predictable, our processing of that environment should show increasing structure. In this study, we capture these unfolding processes using the dynamics of arm movements and reveal explicit patterns of predictive arm placement during learning. We tracked the arm using the Nintendo Wii remote while participants engaged in a visuospatial statistical learning task. Over training, the arm exhibits predictive movements towards event sequences that have higher transitional probabilities. With only a short period of training in few subjects, these anticipatory movements reflect the emergence of statistical structure in a visual learning space.

Keywords: statistical learning; action; dynamics.

Introduction: Statistical Learning

Whether we intend to or not, we are quite good at integrating statistically probable events into meaningful patterns. These patterns are learned quickly and automatically despite minimal regularities in an input environment (e.g., Aslin, Saffran, & Newport, 1998; Cleeremans, Destrebecqz, & Boyer, 1998; Conway & Christiansen, 2005; Saffran, Aslin, & Newport, 1996; Perruchet & Pacton, 2006). One well-known demonstration of statistical learning is the segmentation of words from a continuous speech stream (e.g., Saffran et al., 1996). In these studies, participants listen to different syllables that steadily repeat in seemingly random order. Unbeknownst to them, the stream contains sequences of syllables that progress together in a particular order. These sequences, or "words," are presented randomly to obscure any perceptual boundaries that might readily distinguish them. Instead, the critical information needed to detect the word boundaries is transitional probabilities that are higher for syllables within the word than for syllables between the words. After extended exposure, humans are able to effectively use the transitional probabilities to parse the speech stream.

These findings are analogous to those in other perceptual domains, most notably visual and scene perception (e.g., Cleeremans et al., 1998; Kirkham, Slemmer, & Johnson, 2002). Much like the structure in language, a visual landscape contains a complex configuration of probabilistic features. However, the features also unfold on both temporal and spatial dimensions. Still, humans readily make use of transitional probabilities that occur within these changing scenes. For example, adults can easily learn the subtle statistical patterns that occur across a continuous sequence of moving shapes (Fiser & Aslin, 2002). There is also a strand of research that incorporates motor responses into the learning process (e.g., Hunt & Aslin, 2001). To do so, a serial reaction time paradigm is used to record button presses as visual targets appear in one of several locations on a computer screen. Although the participants may not consciously be aware of the statistical regularities, reaction time decreases between sequences of targets that occur with a higher probability, thus suggesting that the cognitive system is anticipating target occurrences. The advantage of this method over passive listening or viewing of stimuli is that learning can be tracked in real time.

In these and other cases, the change in anticipatory strength is inferred through a measurement of reaction time. While a strong inference, it only indirectly captures the perceptual changes necessary to learn a statistical structure that permits anticipatory, predictive processes. To reveal such processes more directly, we turn to the tracking of action dynamics.

Current Study: Seeking Explicit Prediction

One common proposal for statistical learning processes is prediction, since it serves as "self-corrective" feedback as a system processes events in time (Cleeremans et al., 1998; Elman, 1990). In most studies of statistical learning, this predictive process is an inference from reaction time data – responses are facilitated, but no explicitly predictive behavior is measured. Indeed, some have recently challenged the centrality of prediction by seeking to explicitly measure it (e.g., Jones & Pashler, 2007).

Here, we also seek an explicit behavioral measure of predictive processes. With action dynamics, continuous movements are viewed as a direct reflection of cognitive activity (Spivey & Dale, 2006). Indeed, the mind and body are known to covary in a multitude of cognitive tasks, ranging from low-level speech perception (Spivey, Grosjean, & Knoblich, 2005) to paired-associate learning over longer time-scales (Dale, Roche, Snyder, & McCall, 2008). In these experiments, the arm is tracked as participants engage in various decision and learning tasks. The resulting movements provide insight into the dynamics of underlying cognitive processes.

In the current study we use the arm to track the emergence of visual statistical learning. By doing so, we are able to capture predictive movements between "syllable" targets in a visual space. In this task, syllables correspond to fixed regions on a planar surface, and "words" are represented as a unique path between three of the syllable locations. Participants are instructed to click on each visual cue as soon as it appears in the sequence. As participants use their arm to follow cues through this space of positions, they are unwittingly exposed to the statistical pattern of a simple artificial language. This pattern is similar to previous examples where transitional probabilities are higher for syllables within words than for syllables between words.

Crucially, if the probabilities are being learned, the arm will show predictive movements towards the 2nd and 3rd syllable regions after leaving the 1st syllable region. During the period when the visual cue disappears, the learner has the (implicit) opportunity to predict the most likely region of the cues reappearance. By recording the coordinate position of arm movements within this period, we can explicitly show how close the arm gets to the next syllable.

This approach thus achieves a direct and novel measure of anticipation as learning proceeds. In the following two experiments, we track the coordinates of arm movements as participants learn statistical regularities in a visual state space. These data are collected with a handheld Nintendo Wii remote (Wiimote; Dale et al., 2008). Compared to data collection with a computer-mouse, the Wiimote has a notable advantage of increasing the range of arm motion during response performance. In addition, rather than being confined to a desk (a stabilizing surface), the Wiimote is held with an outstretched arm toward the direction of presented stimuli (see Figure 1). This permits a "floating" instability in the arm, during which such explicitly predictive movements may be made towards within-word syllable targets.

Figure 1: Depiction of the experimental context.

Experiment 1: Unmarked Syllable Regions

In this first of two experiments, we presented syllable stimuli on a large blank screen without any conspicuous demarcations of syllable regions. These circle stimuli popped in and out of spatial locations governed by an artificial grammar.

Participants

Nine undergraduate students participated in exchange for course extra credit.

Stimuli

Sequences occurred across 12 spatially distinct regions in a visual space. The regions were arranged in a 3 x 4 array that was not visible to participants. During sequence presentation, a large black circle appeared in one of the 12 regions. The circles followed a spatial pattern that allowed for language-like statistical regularities."Word" corresponded to a specific configuration of three different spatial regions. A total of four words were created, such that no word contained syllable regions that were used by another word. The syllable regions for each word were also selected to be equidistant. These distance relations were maintained across the four words.

To create a syllable sequencing that was similar to the pseudo-language of Saffran et al. (1996), the word configurations were randomly concatenated with the restriction that exact word configurations could not immediately follow each other. The result was a string of 1296 syllable presentations, with each word configuration occurring 108 times (i.e., 108 occurrences * 4 words * 3 syllables per word). As shown in Figure 2, syllables 2 and 3 are predicted with a probably of 1 following any syllable 1. Syllable 1, however, is only .33 probable given the completion of any individual word.

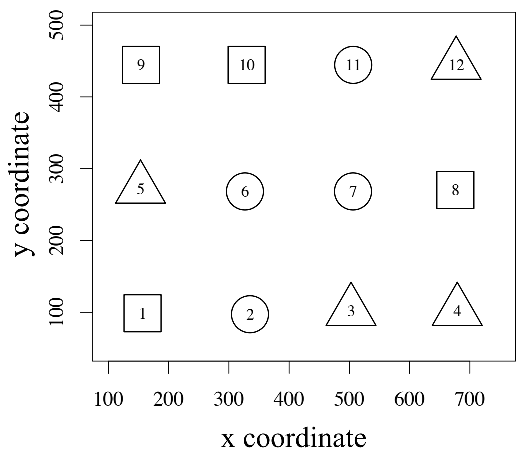

Figure 2: The "syllable" locations on the screen. Symbols to denote syllable types here are used throughout the paper below. Syllables serving as word-initial and the least predictable are presented as squares. Syllables 2 and 3 are presented as circles and triangles, respectively. Words are: (9,7,5), (10,2,12), (8,6,4), and (1,11,3).

Procedure

Participants stood about 10 feet in front of a projector screen mounted on a wall. A stimulus circle appeared on the screen and participants were instructed to navigate the handheld Wiimote controller to the circle and click on it. Once they clicked, the screen went blank for 500 ms before the next circle in the syllable sequence appeared. Participants were not told beforehand that there was a statistical pattern hidden amongst the circles.

The entire experiment session was divided into three training blocks, with a two-minute rest between blocks. Participants were encouraged to complete the task as fast as possible. On average, the experiment required 35 minutes. During this time, the x,y coordinates of the Wiimote cursor were automatically recorded and stored for later analysis.

Measures Among the many measures that can be extracted from the unfolding dynamics of the arm during responses, we chose 4. The first 3 are conventionally used in reaction-time and mouse-movement research: total response time, movement initiation time, and time the hand is in motion. A fourth measure was chosen in order to test for the presence of anticipatory movements. We extracted the starting location of the mouse cursor just prior to the presentation of the next syllable. If anticipatory movements are occurring, then the distance to the next target should diminish over training blocks – but only for syllables 2 and 3, the only targets for which predictions can be reliably made. Crucially, this predicts an interaction between syllable type and block. Such an interaction would demonstrate that distance (anticipation) is reducing more for predictable syllables than syllable 1.

Results

Subjects immediately adopted a strategy to facilitate their responding: They moved their cursor to the center of the screen prior to each stimulus (see Figure 3). This was likely to compensate for their initial ignorance of the pattern, and to maintain the cursor in a region equidistant to possible next stimuli. We therefore included the locations of stimuli in the following analysis (since syllables in the second position are closer to the center than syllables 1 and 3). A mixed-effects model, using the SPSS procedure MIXED, was constructed with subject and stimulus location (nested in syllable type to statistically control for the centering tendency) as random factors, and syllable type (1, 2, or 3) and training block (1, 2, or 3) were introduced as fixed factors. We consider the results of each of the four measures individually.

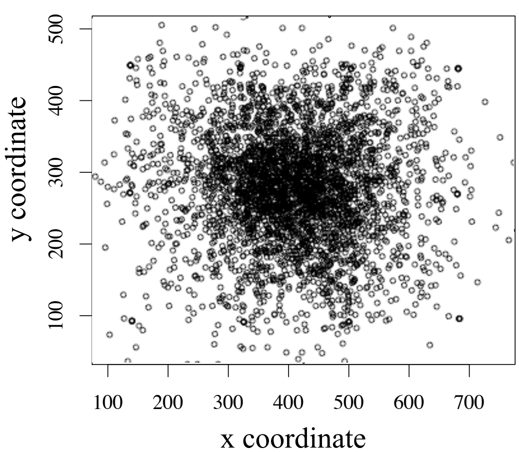

Figure 3: Starting position of all trials for block 1 in Experiment 1. Participants adopted a strategy of centering the cursor prior to the next syllable's presentation.

Total response time. We found a significant main effect for block [$F(2,10780.3) = 97.2, p < .001$], syllable [$F(2,9.0) = 5.3, p < .05$], and a significant block-by-syllable interaction [$F(4,10774.0) = 2.5, p < .05$]. Pairwise comparisons adjusted using Least Significant Difference revealed that all three blocks had significant differences in response time: 663ms, 557ms, and 552ms for blocks 1, 2, and 3 respectively (p's$< .005$). Using the same type of comparison, syllable 2 had the fastest response times, at 513ms. Syllables 1 (655ms) and 3 (624ms) did not differ, but were both significantly longer than syllable 2 (p's$< .05$). The significant interaction appeared to be due to a sharp drop in response time to syllable 2 on the third block.

Movement initiation. There was a significant main effect for block [$F(2,10535.4) = 5.9, p < .005$], and a strong block-by-syllable interaction [$F(4,10530.1) = 7.4, p < .001$]. Syllable only approached significance [$F(2,9.0) = 3.8, p = .07$]. Pairwise comparisons showed that block 3 induced slightly faster initiation times compared to the other blocks: 150ms, 152ms, and 145ms for blocks 1, 2, and 3 respectively (p's$<$ .05). As before, the strong interaction appeared to be due to a sharp drop for syllable 2 on the third block.

Time in motion. Block was significant [$F(2,10536.6) = 5.9, p < .001$], as was syllable [$F(2,9.0) = 4.4, p < .05$], and a weak but significant interaction [$F(4,10529.9) = 2.8, p < .05$]. Syllable only approached significance [$F(2,9.0) = 3.8, p = .07$]. Block 3 had a slightly faster movement times compared to the other blocks: 515ms, 429ms, and 412ms for blocks 1, 2, and 3 respectively (p's$< .05$). Syllable 2 had the fastest movement time (379ms), and differed significantly from syllable 1 (514ms, $p < .05$), but not from syllable 3 (464ms, which did not differ significant from 1).

Distance from starting position to target. Block [$F(2,10780.8) = 117.5, p < .001$], syllable [$F(2,9.0) = 5.7, p < .05$], and the interaction [$F(2,10774.2) = 3.7, p < .01$], were significant. Distance on block 3 (231 pixels) was the smallest compared to block 2 (238.8 pixels) and block 1 (261.9 pixels) (p's$< .001$). Syllable 2 had the closest starting points compared to the 2 syllable types (171 pixels vs. 287.4 and 171.4 pixels for 2, 1, and 3 respectively; p's$< .05$). The interaction in this measure appeared to be due to a substantial drop from blocks 2 to 3 for both predictable syllables, but not for syllable 1 (see Figure 4).

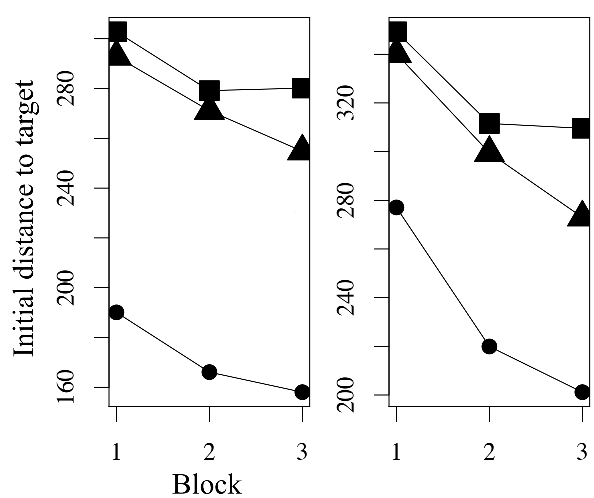

Figure 4: Initial distance to target before the syllable region is presented. As in other figures, squares, circles, and triangles reflect syllable type 1, 2, and 3 respectively. Syllable region types 2 and 3 continue to drop at block 3, while type 1 does not. Left panel: Experiment 1. Right panel: Experiment 2.

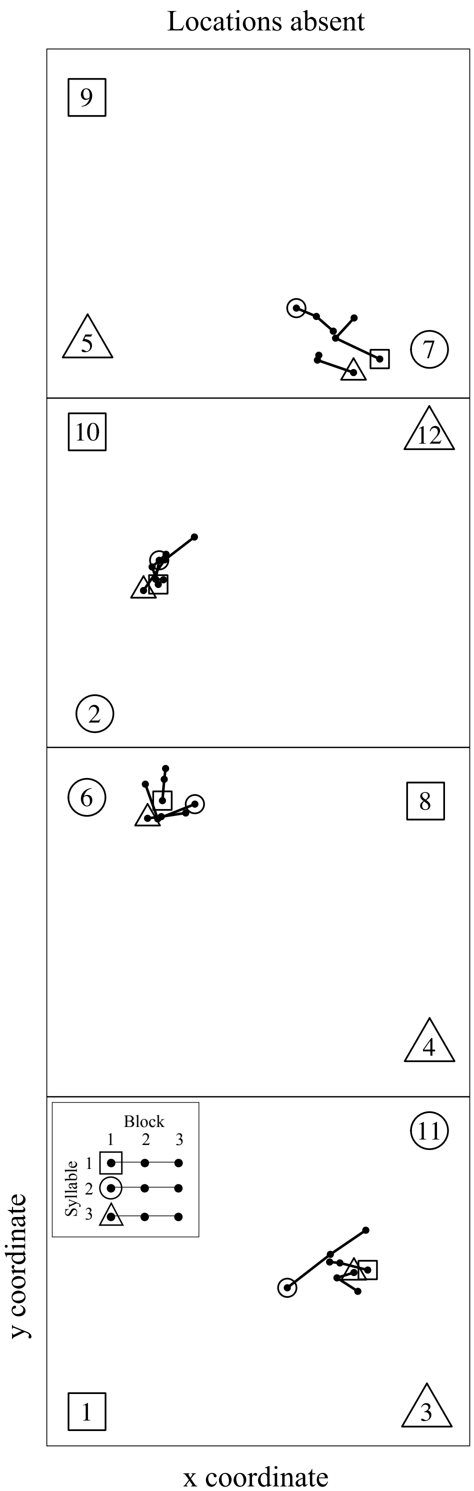

Figure 5: Mean starting positions to each syllable of all four words. Points reflect where a trajectory began immediately prior to the presentation of the stimulus. Data for three syllable types, and three learning blocks, are shown. For example, in the bottom panel (word 1-11-3), participants are gradually initiating their trajectory closer to the syllable 2, before it is presented. Syllable numbers are presented in their locations. For x-y coordinates, see Figure 2.

We plotted the starting positions from block to block for each word (see Figure 5). In this figure, it can be seen that over successive blocks, syllables 2 and 3 induce closer starting positions towards the relevant target, indicating predictive initial movements towards the next syllables, but only for word-internal positions.

Discussion

Participants engaged in the explicit strategy of moving the cursor to the center of the screen. This appears to have enhanced the response patterns to syllables in the second word position, because these are located in the center of the screen. However, the interaction of syllable and block for initial distance indicated that the third syllable also induced a substantial drop over training blocks. This general pattern held across the measures: Syllable 3 responses were generally faster and initially closer to their target than those of syllable 1. Importantly, we factored in syllable position in the visual array as a random factor in our mixed-effects model to ensure that these positions were not solely responsible for the variance in response dynamics. Nevertheless, we designed a simple change to this experimental interface in an aim to reduce this strategy by the participants.

Experiment 2: Demarcated Syllable Regions

Experiment 2 adds a fixed visual array to the background of the visual syllable space. The array serves the purpose of clearly demarcating the 12 syllable regions. This may help avoid having subjects engage in the "center-moving" strategy.

Participants

Nine undergraduate students participated in exchange for course extra credit.

Stimuli and Procedure

The same visual layout was used here as in Experiment 1, except for the presence of the 3 x 4 array that separates and demarcates the 12 syllable regions. As syllables appear in each region, the outline is simply filled in with the stimuli circle, giving the appearance of "buttons" turning on and off, but their locations still visible in the array to the participants.

The syllable sequences that correspond to each word are also the same here as in Experiment 1. However, word configurations were again randomly concatenated with the exception that no two word configurations could occur next to each other. All other aspects of the procedure are identical.

Results

Participants tended to keep their cursor closer to syllable regions than in the previous experiment (see Figure 6). However, even at the first block of training, participants were also engaging in the centering strategy. We subjected the data to the same analysis as described in the previous experiment.

Total response time. The same mixed models were computed for Experiment 2's results. As before, there was a significant main effect for block $[F(2, 11363.5) = 97.0, p <$

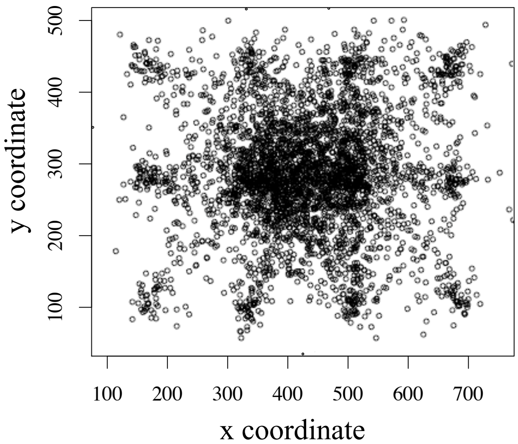

Figure 6: Block 1 starting positions in Experiment 2.

.001] and syllable [$F(2,9.0) = 5.1, p < .05$]. The interaction between them was not significant. Pairwise comparisons revealed the third training block as the one that induced fastest reaction time: 775ms, 627ms, and 602ms for blocks 1, 2, and 3 respectively. Block 3 was significantly faster than 1, but only marginally so compared to 2. Using the same type of comparison, syllable 2 had the fastest response times, at 571ms. Syllables 1 (756ms) and 3 (677ms) did not differ. Syllable 2 was significantly faster than 1 ($p < .05$), but not 3.

Movement initiation. There was a significant main effect for block [$F(2,11181.3) = 48.8, p < .005$], syllable [$F(2,9.0) = 11.2, p < .005$], and a strong block-by-syllable interaction [$F(4,11180.1) = 6.7, p < .001$]. Block 3 induced slightly faster initiation times compared to the other blocks: 190ms, 156ms, and 153ms for blocks 1, 2, and 3 respectively. Blocks 2 and 3 were both significantly faster than the first training block (p's$< .001$). Syllable 2 was again the fastest, at 151ms vs. 158ms and 190ms for syllables 2, 1, and 3 respectively. In fact, syllable 3 was slower than 1 and 2 in terms of initiation time (p's $< .01$). The robust interaction term suggested, in fact, that syllable 3 induced more learning across blocks (a drop of over 62ms) than syllables 1 or 2 (drops of 26ms and 23ms, respectively, in total, from block 1 to 3).

Time in motion. Block was significant [$F(2,11181.6) = 79.0, p < .001$], as was syllable [$F(2,9.0) = 5.0, p < .05$], with no interaction. Syllable only approached significance [$F(2,9.0) = 3.8, p = .07$]. Block 3 again had faster movement times compared to the other blocks: 580ms, 476ms, and 444ms for blocks 1, 2, and 3 respectively (p's$< .01$). Syllable 2 had the fastest movement time (418), and differed significantly from syllable 1 (593, $p < .05$), but not from syllable 3 (489ms, which differed only marginally from 1, $p = .09$).

Distance from starting position to target. Block [$F(2,11362.4) = 413.6, p < .001$], and the interaction [$F(2,11362.0) = 14.3, p < .001$], were highly significant. Syllable only had a marginal effect [$F(2,9.0) = 3.4, p = .08$]. Distance on block 3 (261.2 pixels) was the smallest com-

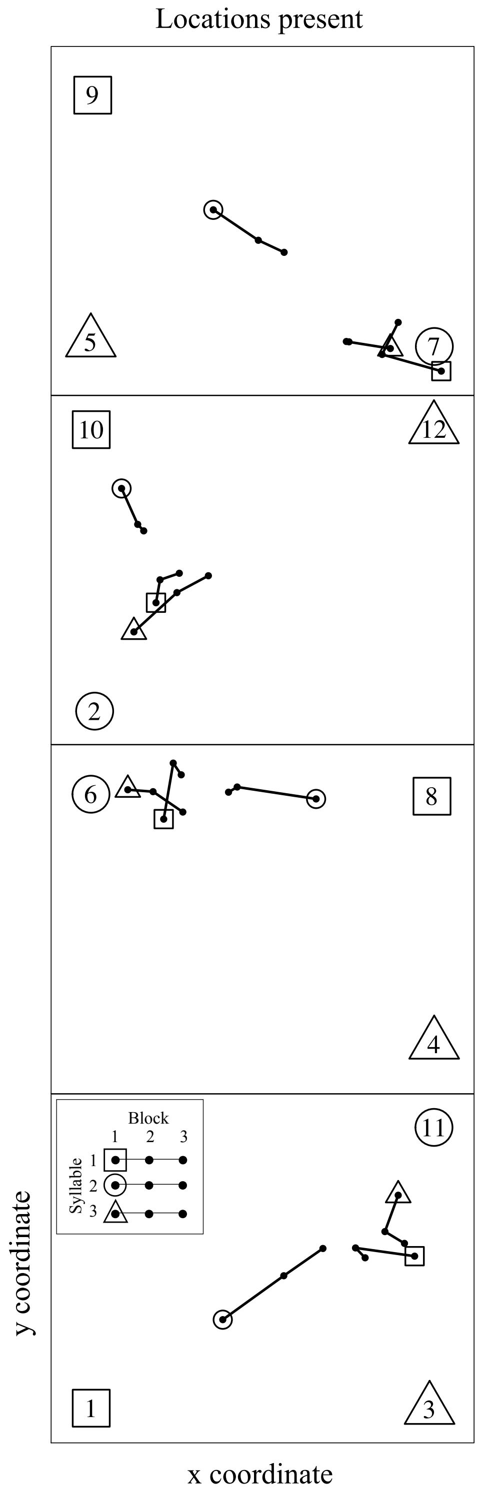

Figure 7: As in Figure 5, points reflect where a trajectory is initiated immediately prior to the presentation of the stimulus. As in Experiment 1, in the panel (word 1-11-3), participants are initiating their trajectory closer to the syllable 2, across learning blocks, before the visual stimulus is presented. Syllable numbers are presented in their locations, with x-y coordinates corresponding to Figure 2.

pared to block 2 (277.0 pixels) and block 1 (322.2 pixels) (p's$< .001$). Syllable 2 had the closest starting points compared to the 2 syllable types (232.7 pixels vs. 323.5 and 304.2 pixels for 2, 1, and 3 respectively; only 1 vs. 1 was significant, $p < .05$). The interaction in this measure again appeared to be due to a substantial drop from blocks 2 to 3 for both predictable syllables, but not for syllable 1 (see Figure 4). As in Experiment 1, we plotted the starting positions from block to block for each word (see Figure 7). Syllables 2 and 3 induce closer starting positions towards the relevant target.

Discussion

This experiment replicates the findings of the previous. Again, the centering strategy produces an enhanced responding pattern to syllable 2. Yet, we controlled for syllable position (as a random factor), and the syllable-type factor remained reliable. The interaction with initial distance is more robust in this experiment, and again shows more anticipatory movements for the predictable syllables. Though the presence of the array did not function to completely remove this curious centering strategy, these results nevertheless are consistent with an enhanced pattern of learning when these "attractors" are visible for participants to anchor to them.

General Discussion

In this study, we investigated statistical learning as the arm moves in a visual state space. The movements followed a pattern of visual stimuli that appeared and disappeared in different regions of space. This pattern had statistical regularities, with words occurring as deterministic sequences between three distinct spatial regions (i.e., trisyllabic words). Overall, the results revealed greater anticipatory movement toward the second and third syllables across the three training blocks. In short: Predictive movements were generated by participants, manifesting the statistical learning. This held true despite the curious centering tendency of our participants, as we statistically controlled for these syllable positions. Through some contest the central role of prediction (Jones & Pashler, 2007), our results suggests that explicit predictive processes are taking place.

We aimed to capture the continuous statistical learning in a visuomotor task using the Nintendo Wiimote, especially well-suited to this goal. Such a method offers the empirical means to move beyond reaction time as a measure of statistical learning, and to begin engaging in a more direct assessment of prediction. However, we do acknowledge that the transitional probabilities used in this study are relatively simple. In addition, using a circular arrayed stimulus setup, more like Hunt and Aslin (2001), may help with the centering issue we encountered in both experiments. A genuine concern is how participants are able to trade-off this responding strategy and genuine statistical learning to predict the next syllable region. In a test phase used in both experiments not reported here (due to space constraints), participants in fact revealed that they encoded the syllable regions in sequence rather than the specific location identity.

The strength of the results is nevertheless surprising given the number of subjects and the brevity of training in comparison to other studies. The predictive processes revealed in the dynamics of arm movements – when permitted to "float about" during statistical learning – may capture emerging sensitivity to statistical structure.

Acknowledgments

This work was supported in part by a National Science Foundation (NSF) Graduate Research Fellowship awarded to the first author, and by NSF Grant BCS-0720322 to the second author. The authors would like to thank Brent Fonville for his assistance in data collection.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-Month-Old infants. *Psychological Science*, *9*(4), 321–324.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39.

Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of Paired-Associate learning. *PLoS ONE*, *3*(3).

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of Higher-Order temporal structure from visual shape sequences. *Learning, Memory*, *28*(3), 458–467.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680.

Jones, J., & Pashler, H. (2007). Is the mind inherently forward looking? comparing prediction and retrodiction. *Psychonomic bulletin & review*, *14*(2), 295–300.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), 35–42.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-Month-Old infants. *Science*, *274*(5294), 1926.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in Real-Time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398.