Research Article

Psychological Science
1–8

DOI: 10.1177/0956797617698226
www.psychologicalscience.org/PS

Rapid Statistical Learning Supporting Word Extraction From Continuous Speech

Laura J. Batterink
Department of Psychology, Northwestern University

Abstract
The identification of words in continuous speech, known as speech segmentation, is a critical early step in language acquisition. This process is partially supported by statistical learning, the ability to extract patterns from the environment. Given that speech segmentation represents a potential bottleneck for language acquisition, patterns in speech may be extracted very rapidly, without extensive exposure. This hypothesis was examined by exposing participants to continuous speech streams composed of novel repeating nonsense words. Learning was measured on-line using a reaction time task. After merely one exposure to an embedded novel word, learners demonstrated significant learning effects, as revealed by faster responses to predictable than to unpredictable syllables. These results demonstrate that learners gained sensitivity to the statistical structure of unfamiliar speech on a very rapid timescale. This ability may play an essential role in early stages of language acquisition, allowing learners to rapidly identify word candidates and "break in" to an unfamiliar language.

Keywords
speech segmentation, statistical learning, language acquisition, reaction time, open data, open materials

Received 11/23/16; Revision accepted 2/15/17

Consider the relatively common experience of overhearing a conversation in a completely unknown foreign language. In contrast to speech in one's native language, which is perceived as a sequence of discrete words, an unfamiliar language is generally heard as a seemingly rapid-fire and continuous stream of phonemes, broken only by silences at the end of utterances. This common perceptual experience illustrates one of the very first challenges faced by language learners: the discovery of word boundaries in continuous speech. Speech consists of a continuous stream of sound, and word onsets are not reliably marked by acoustic cues, such as pauses. Parsing this continuous sequence into word units is a central problem of language acquisition and a prerequisite for acquiring other higher-order aspects of language, such as vocabulary, morphology, and syntax.

An emerging consensus is that this problem may be at least partially solved through *statistical learning*, the process of becoming sensitive to statistical structure in the environment. In spoken language, syllables that occur next to one another within words have higher rates of co-occurrence than syllables that occur next to one another across word boundaries, and becoming sensitive to these co-occurrence properties is one mechanism by which learners may identify words in continuous speech (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). Different mechanisms have been proposed to underlie statistical learning. Early studies of statistical learning assumed that learners' ability to solve segmentation tasks could be attributed to their ability to compute conditional probabilities between co-occurring elements in the input (Saffran, 2001; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran & Wilson, 2003). These computations would then lead to the formation of chunks, or word candidates. An alternative interpretation is that statistical learning is driven directly by the extraction of chunks from the input, which are strengthened or weakened according to the laws governing associative memory; sensitivity to statistical structure

Corresponding Author:
Laura J. Batterink, Northwestern University, Department of Psychology, Cresap Laboratory, Room 212, Evanston, IL 60208
E-mail: lbatterink@northwestern.edu

emerges as a by-product of this process. This explanation is favored by a number of different computational models (French, Addyman, & Mareschal, 2011; Mareschal & French, 2017; Perruchet & Vinter, 1998; Shi, Griffiths, Feldman, & Sanborn, 2010; Thiessen & Pavlik, 2013).

Given that speech segmentation is a prerequisite for acquiring language, an important question concerns how quickly learners become sensitive to patterns in continuous speech. Presumably, it should be advantageous for learners to gain sensitivity to these patterns as rapidly as possible, which would enable them to identify word candidates in speech input and pave the way for later, higher-level stages of language acquisition. Although neither of two main mechanistic accounts of statistical learning (statistical computations and chunking) explicitly addresses the timescale of learning, the process of automatically chunking a candidate word may presumably begin after just a single exposure to the word representation; associative-memory mechanisms may link successive syllables together after a single episode. In contrast, the computation of conditional probabilities would presumably require a lengthier period of exposure, as the learner must gradually accrue information about the statistical properties of the input in order to compute conditional probabilities between different elements.

This question of how quickly statistical learning of speech patterns occurs has not been well addressed. In typical laboratory studies of speech segmentation, learners are exposed to a continuous stream of speech made up of repeating three-syllable nonsense words and later tested to assess the extent of learning. Infants are usually tested with a visual fixation measure, whereas adult testing typically involves a forced-choice recognition task between previously presented items and foils. Speech-segmentation studies using this approach have found evidence of learning in infants after an exposure period of only 2 min (Saffran, Newport, & Aslin, 1996), while longer exposure periods (e.g., 21 min) are more common in studies of older children and adults (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Although these studies suggest that statistical learning of speech patterns can occur relatively quickly, at least in a constrained artificial-language context, this general approach of using an off-line test to measure learning after an arbitrary amount of exposure has not been well suited to investigate the time course of learning.

The goal of the present study was to address how quickly learners become sensitive to patterns in continuous streams of speech. Following previous studies of speech segmentation, I exposed participants to continuous auditory streams of repeating trisyllabic nonsense words, without any pauses or other auditory cues marking word boundaries. However, in contrast to most other studies, the present work used an on-line measure of statistical learning based on reaction time (RT), which required participants to respond to target syllables. This target-detection task has been previously shown to be sensitive to statistical learning, as reflected by faster RTs to predictable than to unpredictable syllables occurring at the beginnings of words (Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015; Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015). Each syllable stream in the present study was composed of a novel set of repeating nonsense words, which ensured that statistical learning began from square one for each stream. I hypothesized that RT effects indexing learning would emerge within several exposures to a novel word. This finding would provide evidence that learners become sensitive to statistical patterns in speech very rapidly, a process that facilitates the identification of words and ultimately the acquisition of other aspects of language.

Method

Participants

A total of 19 young English-speaking adults (11 women, 8 men; mean age = 20.2 years, *SD* = 1.7) participated in this study. I originally targeted a sample size of 15 to 20 participants on the basis of results from a previous behavioral study conducted at Northwestern University's Cognitive Neuroscience Lab, which used a similar RT task to measure statistical learning (Batterink, Reber, Neville, & Paller, 2015). Two separate groups, each with 12 participants, exhibited large and robust learning effects with a post hoc power of 99%. The present design included a larger number of trials per participant and condition (48 trials within each Word Presentation × Triplet Position bin) relative to the original study (36 trials per triplet position), which further increased power. Thus, I expected that a sample size of 15 to 20 participants should be more than adequate to reveal learning effects. I planned to complete data collection within a single academic term, provided that data were collected from at least 15 participants, and would terminate data collection after reaching a sample size of 20. At the end of the academic term, I had successfully run 19 participants. All of the procedures and protocols followed the guidelines of the Northwestern University Institutional Review Board.

Stimuli

Two syllable inventories, each consisting of 24 unique syllables, were constructed. One syllable inventory was recorded by a male native-English speaker and the other by a female native-English speaker, both using neutral intonation. Individual sound files, each containing a single syllable, were created from the recordings. The beginning of each sound file coincided with the precise onset of the syllable. All sound files had an approximate duration of 220 to 250 ms and were equated for perceived

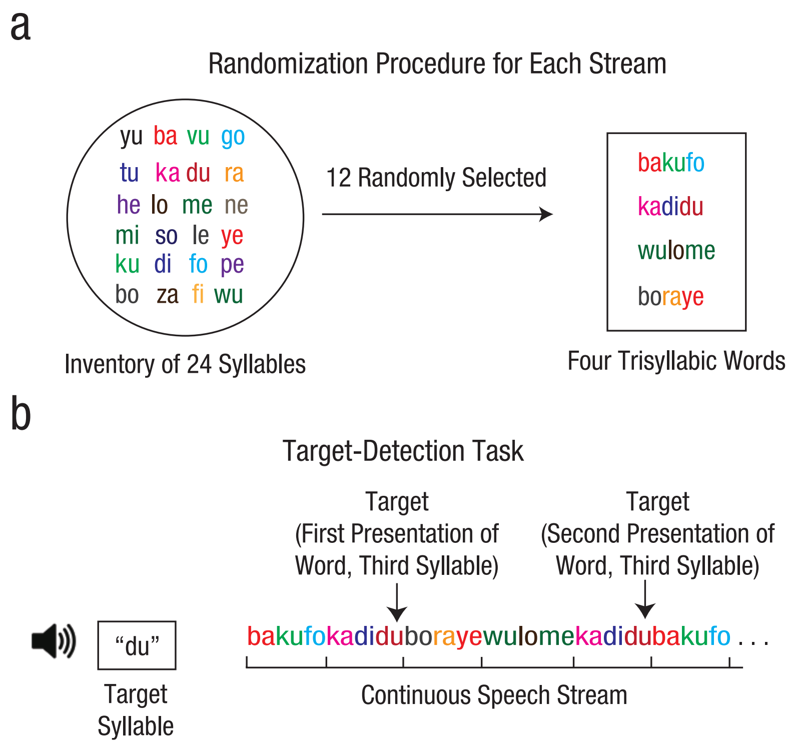

Fig. 1. Stimulus design (a) and example trial sequence (b). Each speech stream was composed of a randomly selected subset of 12 syllables drawn without replacement from a pool of 24 syllables. These 12 syllables were then randomly distributed to create four trisyllabic words, which were repeated four times each. The resulting 16 words (48 syllables) were concatenated together in pseudorandom order and presented aurally without any pauses between them. Before each stream was played, participants saw and heard a target syllable and were asked to identify it as quickly as possible every time it occurred in the stream.

volume. Continuous speech streams were created by concatenating the individual syllables together in a preset order, at a rate of 300 ms per syllable.

Procedure

For each syllable stream, a random subset of 12 syllables was drawn without replacement from the pool of 24 possible syllables in each syllable inventory and randomly distributed to create four different trisyllabic "words" (Fig. 1a). This unique set of four repeating words allowed learning to be measured on a very short timescale. Each speech stream consisted of the four words repeated four times each; the resulting 16 words (48 syllables) were concatenated together in pseudorandom order, with the constraint that the same word did not repeat twice in a row. A specific syllable served as the target for each stream. The target syllable never occurred at the first or last two positions of the syllable stream. Each of the 24 syllables of the syllable inventory served as the target syllable three times, for a total of 72 streams for each voice (male and female). Participants thus listened to a total of 144 streams. Voice order (male first or female first) was counterbalanced across participants. For both voices, the number of targets in each triplet position (first, second, or third syllable within a word) ranged from 45 to 53 per participant. Each of the 24 syllables was represented an equal number of times across all streams.

At the beginning of each trial, participants were presented with the written target syllable (e.g., "du") and an auditory sample of the target. The written syllable then remained on screen while participants listened to the stimulus stream (Fig. 1b). Participants were instructed to respond to each target syllable as quickly and accurately as possible. If statistical learning occurred during the individual streams, it was expected that RTs would be fastest to targets that occurred in the final position of a word, with targets occurring at the beginning of a word and targets occurring in the middle of a word eliciting the slowest and intermediate RTs, respectively. These effects were expected to require at least one exposure to the word, emerging sometime between the second and fourth word presentation.

Data analysis

Robust linear mixed-effects modeling was used to account for repeated measures. RTs to targets ("hits") were measured at the individual trial level for each participant and classified according to the following factors: participant (1–19), word presentation (1st, 2nd, 3rd, or 4th occurrence of the word in the stream), triplet position (1st, 2nd, or 3rd syllable in the word), and stream position (3rd through 46th syllable in the stream; targets never occurred at the first or last two positions of the stream). Model fixed effects consisted of word presentation, triplet position, overall stream position, and the interaction between word presentation and triplet position. To select random effects, I used the method of Bayesian information criterion (BIC) penalized likelihood.

In the initial full model, participant was included as a random intercept, and random slopes for participants were included for all fixed effects. BIC values were computed for the initial full model and other alternative models that included one or more random slopes for the different fixed factors in all possible combinations. The final best model (associated with the lowest BIC value) included participant as a random intercept and stream position as a random slope. Random slopes for the remaining factors (word presentation, triplet position, and the interaction between word presentation and triplet position) were not significant and resulted in higher BIC values, and thus were excluded from the final model. Word presentation was modeled as a categorical predictor variable, because there is a categorical difference between the first presentation of a word (prior to any opportunity for learning) and subsequent presentations. Stream position and triplet position were modeled as continuous predictors, because both variables were originally conceptualized as continuous and were empirically found to show significant linear relationships with RT in exploratory regression analyses—stream position: $F(1, 9561) = 28.4$, $p < .001$; triplet position (excluding Word Presentation 1): $F(1, 7092) = 64.7$, $p < .001$. Continuous predictors were centered such that the intercept represented the first value for both variables (stream position = 3, triplet position = 1). Stream position was not a variable of direct interest but was included as a predictor in the model in order to control for possible influences of this effect.

The central hypothesis of the study was whether RT effects indexing statistical learning would emerge within several exposures to a novel word; this was tested by examining the interaction between word presentation and triplet position. A significant interaction between these two factors was characterized through the parameter estimates of the RT slope within each word-presentation condition and through follow-up analyses that tested whether the RT slope within each word-presentation condition was significantly different from zero. These follow-up analyses examined when the earliest evidence of significant priming emerged, the main question of interest, and were conducted separately within each word-presentation condition using the same predictors as in the original RT model. An additional follow-up analysis was conducted that compared the RT slope estimates between each of the first three word presentations in a stream and the subsequent presentation (i.e., $n < n + 1$), in order to examine whether learning effects followed an expected learning curve, gradually increasing as a function of exposure to the underlying words.

In addition, for each triplet position, pairwise comparisons between the first word presentation (representing the baseline condition) and subsequent word presentations were conducted in order to determine whether RTs were faster in response to targets in the predictable syllables of words (Triplet Positions 2 and 3) and slower to targets in the unpredictable syllables (Triplet Position 1). Such RT differences have been previously shown to result from statistical learning (Turk-Browne, Scholl, Johnson, & Chun, 2010). These comparisons were computed on model estimates of the mean of each word presentation at each triplet position, evaluated at the first stream position for targets (i.e., the third overall position in the stream). Bonferroni corrections at the level of each triplet position were applied to these pairwise comparisons (i.e., $p = .05/3$ pairwise comparisons).

As described in Results, one unexpected finding was that targets that occurred later in the stream elicited significantly slower RTs relative to targets that occurred earlier in the stream. In order to exclude the impact of stream position while visualizing the main factors of interest (i.e., word presentation and triplet position), I also plotted the RTs for each triplet-position-by-word-presentation cell using model estimates for word presentation, triplet position, and their interaction. This additional plot was included simply as a way to more clearly visualize the main effects of interest, relative to a plot of raw mean RTs that also reflected the effect of stream position.

Paralleling the main RT analysis, a follow-up analysis examined whether target detection was influenced by word presentation and triplet position. Targets were coded as "detected" only if they were followed by a response less than 1,200 ms after presentation and were otherwise coded as "missed." I conducted a mixed-effects logistic regression model with the same factors as in the RT model (participant, word presentation, triplet position, and stream position). Of all these factors, only participant was ultimately included as a random effect, as random effects for the remaining factors were not significant and decreased model fit.

Across all participants, a total of 10,944 trials were available for analysis. Only detected targets were included in RT analyses. All of the analyses were conducted using SPSS statistical-analysis software. All p values are from two-tailed tests with an alpha of .05.

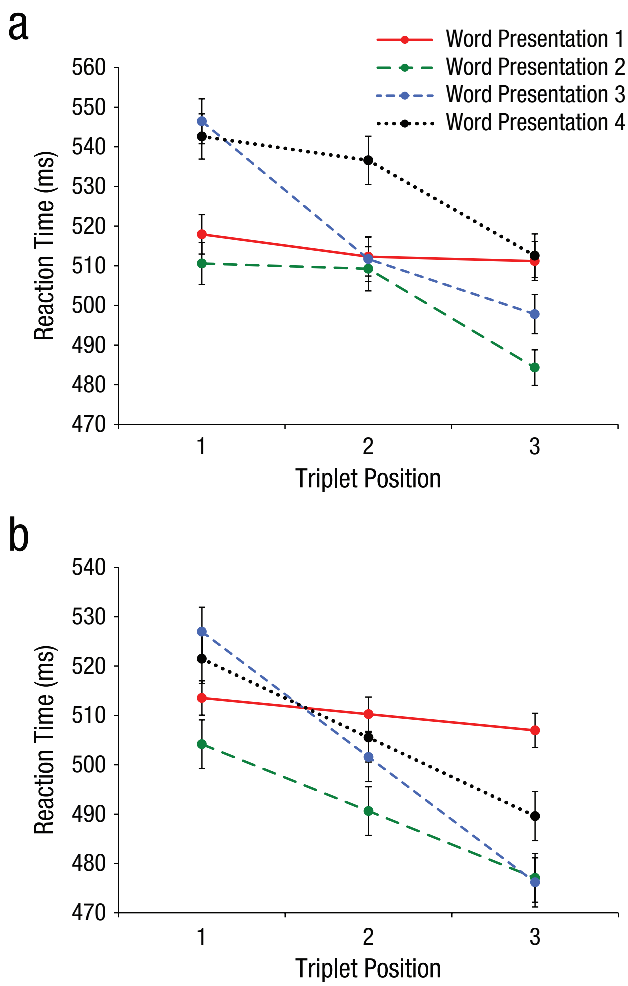

Fig. 2. Reaction time (RT) data as a function of triplet position (first, second, or third syllable in the word) and word presentation (first, second, third, or fourth occurrence of the word in the stream). Overall mean RT is shown in (a). Mean RT predicted by main effects of interest (triplet position, word presentation, and their interaction), controlling for the effect of overall syllable position in the stream, is shown in (b). Error bars represent ±1 *SEM*.

Results

Participants detected an average of 87.4% of targets. Mean RT as a function of triplet position and word presentation is shown in Figure 2a. Mean RT predicted by model estimates of the main effects of interest—word presentation, triplet position, and their interaction—are shown in Figure 2b. As described in Data Analysis, Figure 2b is provided to clearly present only those effects that are relevant to the main hypothesis, as data shown in Figure 2a reflect the additional effect of stream position, a confounding variable not of interest in the present study. This follow-up analysis was necessary because targets that occurred later in the stream elicited significantly slower RTs relative to targets that occurred earlier in the stream, $F(1, 160) = 5.41$, $p = .021$; stream-position coefficient = 0.72 ms, SE = 0.31, 95% confidence interval (CI) = [0.11, 1.33]; log-likelihood ratio = 21.1.

Consistent with my hypothesis, results showed that RTs to detected targets over the three triplet positions differed significantly as a function of word presentation, $F(3, 9527) = 6.68$, $p < .001$; log-likelihood ratio = 20.0. Estimates of fixed effects indicated that at the first word presentation, RTs across the three triplet positions were stable (triplet-position coefficient = −3.35 ms, SE = 3.47, 95% CI = [−10.14, 3.45]). However, by the second word presentation, a robust RT effect was already present; targets that appeared later in a word (and were thus more predictable) elicited faster RTs than targets at the beginning of a word (triplet-position coefficient = −13.6 ms, SE = 4.91, 95% CI = [−23.20, −3.94]). This effect was also present for targets occurring at both the third word presentation (triplet-position coefficient = −25.5 ms, SE = 4.97, 95% CI = [−35.20, −15.70]) and the fourth word presentation (triplet-position coefficient = −16.0 ms, SE = 4.98, 95% CI = [−25.70, −6.19]). Follow-up analyses confirmed that the RT slope across triplet positions was not significantly different from zero at the first word presentation, $F(1, 2451) = 2.13$, $p = .15$; log-likelihood ratio = 2.12, but was highly significant at all subsequent word presentations—Presentation 2: $F(1, 2397) = 16.1$, $p < .001$, log-likelihood ratio = 16.0; Presentation 3: $F(1, 2342) = 50.5$, $p < .001$, log-likelihood ratio = 49.9; Presentation 4: $F(1, 2300) = 17.1$, $p < .001$, log-likelihood ratio = 17.1. Thus, RTs were faster to more predictable syllables after only a single word presentation, which provides evidence of rapid statistical learning of sound patterns in continuous speech.

Next, I examined whether the magnitude of the learning effect (i.e., the negative RT slope across triplet positions) increased with additional word presentations, following an expected learning curve. I found partial support for this idea. The effect of triplet position was significantly larger for the second word presentation compared with the first word presentation (triplet-position effect = 10.2 ms, SE = 4.91, 95% CI = [0.59, 19.90]), $t(9526) = 2.08$, $p = .038$, and for the third compared with the second word presentation (triplet-position effect = 11.9 ms, SE = 4.98, 95% CI = [2.12, 21.70]), $t(9526) = 2.39$, $p = .017$. However, the effect of triplet position did not significantly increase from the third to the fourth word presentation, but rather became marginally reduced (triplet-position effect = −9.51 ms, SE = 5.05, 95% CI = [−19.40, 0.38]), $t(9526) = -1.88$, $p = .060$. In sum, the learning effect increased as a function of exposure after several word presentations, but did not continue to grow from the third to the fourth word presentation.

In principle, the learning effect could reflect facilitation to predictable syllables (Triplet Positions 2 and 3) or a delay to unpredictable syllables (Triplet Position 1), as has been previously demonstrated to result from statistical learning (Turk-Browne et al., 2010). Pairwise comparisons on the estimated marginal means of word presentation within each triplet position supported only the former possibility. For unpredictable syllables (i.e., those that occurred at the start of words), no significant differences were found between the first presentation of a word and subsequent presentations of a word—Word Presentation 2 – Word Presentation 1: mean difference = −9.43 ms, SE = 7.12, 95% CI = [−26.50, 7.62], t(9527) = 9.43, p > .250; Word Presentation 3 – Word Presentation 1: mean difference = 13.4 ms, SE = 8.60, 95% CI = [−7.18, 34.00], t(95276) = 1.56, p > .250; Word Presentation 4 – Word Presentation 1: mean difference = 7.84 ms, SE = 10.80, 95% CI = [−18.10, 33.80], t(9527) = 0.72, p > .250. Thus, although there is a visual hint in Figure 2b that RTs may be delayed to Word Presentations 3 and 4 relative to Word Presentation 1, consistent with anticipatory effects that have been reported previously (Turk-Browne et al., 2010), these differences were not statistically significant.

When target syllables occurred in the middle of words, RTs to the initial word were significantly different from RTs to the second word, although not from RTs to the third or fourth word—Word Presentation 2 – Word Presentation 1: mean difference = −19.7 ms, SE = 5.00, 95% CI = [−31.60, −7.68], t(9527) = 3.53, p < .001; Word Presentation 3 – Word Presentation 1: mean difference = −8.71 ms, SE = 7.03, 95% CI = [−25.50, 8.13], t(9526) = 1.24, p > .250; Word Presentation 4 – Word Presentation 1: mean difference = −4.77 ms, SE = 9.60, 95% CI = [−27.70, 18.20], t(9526) = 0.50, p > .250. Finally, when target syllables occurred at the end of words, RTs to the initial word were significantly different from RTs to the second and third word but the difference between RTs to the first and the fourth word did not reach significance—Word Presentation 2 – Word Presentation 1: mean difference = −29.9 ms, SE = 6.90, 95% CI = [−46.40, −13.40], t(9527) = 4.33, p < .001; Word Presentation 3 – Word Presentation 1: mean difference = −30.8 ms, SE = 8.63, 95% CI = [−51.40, −10.10], t(9527) = 3.57, p = .001; Word Presentation 4 – Word Presentation 1: mean difference = −17.4 ms, SE = 10.80, 95% CI = [−43.20, 8.43], t(9527) = 1.61, p > .250.

A follow-up contrast confirmed this overall pattern, demonstrating that predictable syllables (Triplet Positions 2 and 3) occurring in later word presentations (2–4) elicited significantly faster RTs overall relative to predictable syllables occurring within the first word (Word Presentation 1 – Word Presentations 2–4: mean difference = 21.3 ms, SE = 8.28, 95% CI = [5.06, 37.50]); t(6435) = 2.57, p = .010. In sum, learning primarily resulted in an overall enhancement in processing more predictable syllables rather than a delay in processing less predictable syllables at the beginning of words.

Finally, I examined whether detection rate differed significantly as a function of word presentation and triplet position (this paralleled the main RT analysis). In contrast to the observed RT effect, detection rate over the three triplet positions did not differ significantly as a function of word presentation, F(3, 10935) = 0.32, p > .250; log-likelihood ratio = −22.9. Thus, unlike RT, detection rate was not a reliable index of statistical learning. The finding that detection did not change as a function of triplet position and word presentation provides evidence that RT differences among conditions did not simply reflect a speed/accuracy trade-off; rather, faster RTs appeared to reflect true facilitation in processing.

Discussion

The results of the present study demonstrate that statistical learning of sound patterns in continuous speech can occur incredibly rapidly. After only a single exposure to the hidden component words of continuous nonsense speech, learners' RTs were faster to more predictable syllables. This RT pattern demonstrates that learners quickly gained sensitivity to the statistical structure of the speech stream and made use of this knowledge during on-line processing, facilitating performance on the task.

The finding that some degree of learning occurred after just a single word exposure suggests that learning was primarily driven by the extraction of chunks from the input, rather than through the computation of conditional probabilities. Logically, the computation of conditional probabilities depends on accruing statistical data across a sample of input and cannot occur instantly after only a single exposure to an underlying pattern. In contrast, evidence of learning after only a single word repetition may be explained by an automatic chunking mechanism. The idea that chunking, driven by associative-memory mechanisms, can give rise to sensitivity to statistical structure is supported by computational models. For example, according to the PARSER model (Perruchet & Vinter, 1998), chunks are formed from a sequence of elements on a random basis, as a natural consequence of the capacity-limited attentional processing of the incoming information. These chunks are then stored in memory and strengthened or weakened according to the laws governing associative memory. If a chunk is encountered again, the activation of its representation increases; otherwise, its representation decays over time. If an element within a chunk occurs in a different chunk, the previously stored chunk is subject to interference, which decreases its activation level. Over time, chunks that form statistically coherent elements within a sequence (i.e., a word) will be strengthened, while chunks with lower probabilities of co-occurrence (i.e., syllables spanning word boundaries) will be forgotten (cf. Thiessen, 2017). Thus, in the present study, the very first exposure to a word might have sometimes resulted in the formation of a chunk, whose representation could then be stored in memory (albeit at a weak level). When the word was subsequently presented, participants may have retrieved the stored representation, which allowed them to anticipate predictable syllables and thus respond more quickly to second- and final-position targets.

Such a chunking mechanism would support rapid word learning, even prior to the emergence of conditional probability computations. Rapid automatic chunking would also specifically allow language learners to take advantage of word repetitions that commonly occur in natural language, a possibility that aligns with previous evidence that word repetition facilitates language learning. For example, infant-directed speech is characterized by the frequent repetition of words, compared with non-infant-directed speech (Cockcroft, 2002), and repetitiveness in maternal input at the age of 7 months predicts language outcomes at the age of 2 years (Newman, Rowe, & Bernstein Ratner, 2016). Another study found that word learning in 2-year-old children was successful only when the names of novel objects were repeated across successive sentences rather than distributed throughout labeling episodes, which suggests that immediate opportunities to detect recurring structure facilitate young children's word learning (Schwab & Lew-Williams, 2016). Given evidence that statistical learning operates in both children and adults (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, et al., 1997), frequent word repetitions may also support second-language acquisition in adult learners.

The chunking account of statistical learning would predict that the magnitude of the learning effect (i.e., faster RTs to predictable than to unpredictable syllables within a word) would increase with additional exposure to the underlying words, reflecting an increase in chunk activation. A comparison of the RT effects between each word presentation and the subsequent one provides partial support for this idea, demonstrating that the learning effect gradually increased from the first to the third word presentation. However, the learning effect did not significantly increase from the third to the fourth word presentation, inconsistent with chunking models. One possible explanation is that the lack of difference between the third and fourth presentations represents nothing more than a statistical blip, rather than the beginning of a long-term trend. Alternatively, other cognitive factors beyond statistical learning may have influenced RTs to the final word presentation. For example, participants may have become aware that each stream contained exactly four targets, and after detecting three targets may have become more cautious or hesitant to respond to the fourth and final target, knowing that only a single target remained. This hesitancy could possibly have led to a slight reduction in the learning effect to the fourth word presentation. By incorporating a design with more than four exposures to the underlying words, future work may test whether the learning effect continues to generally increase with additional exposure to the underlying words, as would be predicted by a chunking account of statistical learning, or whether the effect quickly reaches asymptote after several word presentations.

Finally, one unexpected finding was that RTs were slower for targets occurring later in the syllable stream, an effect that was independent of the number of word repetitions. I suggest that this deterioration in performance over the course of the stream may be due to sensory interference or overload induced by the rapid presentation of previous syllables. Nonetheless, by covarying out effects of stream position, I was able to isolate the effect of word presentation per se and to directly assess effects of statistical learning on performance.

In sum, these results demonstrate that statistical learning of sound patterns in speech operates on a very rapid timescale. The speed with which learning occurs suggests that the automatic chunking of segments from input may be a major mechanism contributing to this type of learning. The efficiency of this mechanism may play a critical role in early stages of language acquisition, allowing language learners to quickly "break in" to an unfamiliar language and paving the way for the acquisition of more advanced components of language, such as semantics and syntax.

Action Editor

Matthew A. Goldrick served as action editor for this article.

Author Contributions

L. J. Batterink is the sole author of this article and is responsible for its content.

Acknowledgments

I would like to thank Kelsey Aaronson for her help in data collection and Ken A. Paller for his comments on an early draft of this manuscript.

Declaration of Conflicting Interests

The author declared that she had no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by National Institutes of Health Grants T32 NS 047987 and F32 HD 078223.

Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/z69fs. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617698226. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78.

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning & Memory*, *22*, 544–556.

Cockcroft, K. (2002). Language development. In D. Hook, J. Watts, & K. Cockcroft (Eds.), *Developmental psychology* (pp. 218–232). Lansdowne, South Africa: UCT Press.

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation segmentation: A new measure of statistical learning in speech segmentation. *Experimental Psychology*, *62*, 346–351.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for the sequence segmentation and chunk extraction. *Psychological Review*, *118*, 614–636.

Mareschal, D., & French, R. M. (2017). TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), Article 20160057. doi:10.1098/rstb.2016.0057

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*, 1158–1173.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246–263.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493–515.

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

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*, 273–284.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*, 879–886.

Shi, L., Griffiths, T., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, *17*, 443–464.

Thiessen, E. D. (2017). What's statistical about learning? Insights from modeling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), Article 20160056. doi:10.1098/rstb.2016.0056

Thiessen, E. D., & Pavlik, P. I. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, *37*, 310–343.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*, 11177–11187.