

Statistical Learning of Melodic Patterns Influences the Brain's Response to Wrong Notes

Toviah Moldwin^{1,2}, Odelia Schwartz^{1,3}, and Elyse S. Sussman¹

Abstract

■ The theory of statistical learning has been influential in providing a framework for how humans learn to segment patterns of regularities from continuous sensory inputs, such as speech and music. This form of learning is based on statistical cues and is thought to underlie the ability to learn to segment patterns of regularities from continuous sensory inputs, such as the transition probabilities in speech and music. However, the connection between statistical learning and brain measurements is not well understood. Here we focus on ERPs in the context of tone sequences that contain statistically cohesive melodic patterns. We hypothesized that implicit learning of statistical regularities would influence what was held in auditory working memory. We predicted that a wrong note occurring within a cohesive pat-

tern (within-pattern deviant) would lead to a significantly larger brain signal than a wrong note occurring between cohesive patterns (between-pattern deviant), even though both deviant types were equally likely to occur with respect to the global tone sequence. We discuss this prediction within a simple Markov model framework that learns the transition probability regularities within the tone sequence. Results show that signal strength was stronger when cohesive patterns were violated and demonstrate that the transitional probability of the sequence influences the memory basis for melodic patterns. Our results thus characterize how informational units are stored in auditory memory trace for deviance detection and provide new evidence about how the brain organizes sequential sound input that is useful for perception. ■

INTRODUCTION

Humans are remarkable pattern learners. Yet it is still a puzzle how we extract meaning from sound. One example is that passive exposure to the spoken speech stream, starting in utero, has been shown to influence language acquisition skills (Gervain & Werker, 2013; Moon, Lagercrantz, & Kuhl, 2013; Vouloumanos & Werker, 2004; DeCasper & Spence, 1986), aiding the segmentation of a continuous speech stream into discrete words (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009; Mattys & Jusczyk, 2001; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Aslin, & Newport, 1996; Friederici & Wessels, 1993). An influential theory suggests that segmentation of the speech signal, which is needed to identify word boundaries, is facilitated by implicit calculations of stimulus statistics (i.e., by passive exposure to the stimulus input; Saffran et al., 1999). This has been demonstrated by a higher transitional probability of within-word phoneme sequences playing a facilitative role in the learning of word boundaries in infants and adults (Aslin & Newport, 2012; Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999; Saffran et al., 1996, 1999; Myers et al., 1996; Friederici & Wessels, 1993; Jusczyk, Cutler, & Redanz, 1993).

It is still not clear, however, how pattern learning through passive exposure to stimulus statistics are instantiated in memory representations and used to perceive events in the environment. Despite the wealth of behavioral experiments pertaining to statistics-based pattern recognition, brain measurements that index statistical learning are not fully understood.

The current study investigated statistical learning of unfamiliar musical phrases. The goal was to assess how sequentially presented sounds would be stored as informational units (melodic sound patterns) in auditory memory. Previous studies investigating transitional probabilities have focused on the response to the deviant—the strength of the response to the deviant depends on the degree of violation of the prediction (Koelsch, Busch, Jentschke, & Rohrmeier, 2016; Daltrozzo & Conway, 2014; Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012; Furl et al., 2011; Mill, Coath, Wennekers, & Denham, 2011; Garrido, Kilner, Stephan, & Friston, 2009). In the current study, we did not manipulate the probability of the deviant but rather used equiprobable deviants. That is, deviants equally violated predictions from the overall sequence. This allowed us to assess the strength of the representation of the melodic phrases in auditory memory based on the response to the deviant. We presented listeners with a tone sequence containing two different musical patterns while brain wave activity was recorded. We used unfamiliar tone patterns so that listeners could

¹Albert Einstein College of Medicine, Bronx, NY, ²The Hebrew University of Jerusalem, ³University of Miami

not rely on prior knowledge of songs or musical advertisements, and participants were not told anything about the patterns in the tone sequence. Thus, participants could only be able to identify the patterns by making use of the stimulus statistics of the tone sequence. Every so often, the regularity of a sequence was violated by one of two kinds of “wrong” notes (or “deviants”). We tested the hypothesis that implicit learning of statistical regularities in the sound signal would influence what was held in working memory. We predicted that the stimulus statistics would influence the melodic units maintained in transient memory. The regularities would thus be “learned” online and would set up the basis for deviance detection. In turn, deviance detection would be indexed using a component of ERPs called the MMN.

The MMN component is elicited on the basis of regularity learning: maintaining a neural memory trace of repeating sound patterns (“standards”) from which the deviant status of an incoming stimulus is determined (Sussman, Chen, Sussman-Fort, & Dinces, 2014; Näätänen, Astikainen, Ruusuvirta, & Huotilainen, 2010; Sussman, 2007; Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Sussman, Ritter, & Vaughan, 1999). MMN does not require an overt behavioral response to be elicited. Previous work has suggested that implicitly learned tone pattern regularities are strongly represented in memory (Näätänen et al., 2010; Sussman, 2007; van Zuijen, Simoens, Paavilainen, Näätänen, & Tervaniemi, 2006; van Zuijen, Sussman, Winkler, Näätänen, & Tervaniemi, 2004; Tervaniemi, Rytönen, Schroger, Ilmoniemi, & Näätänen, 2001; Saarinen, Paavilainen, Schroger, Tervaniemi, & Näätänen, 1992; Schroger, Näätänen, & Paavilainen, 1992) and that deviance detection leading to elicitation of MMN is predicated on the learned memory of detected sound regularities (Sussman et al., 2014; Teinonen & Huotilainen, 2012; Sussman, 2007; Näätänen et al., 2001). This indicates that MMN elicitation is highly context dependent (Sussman, 2007). Thus, MMN is a good tool to use to study melodic sound pattern learning. We therefore combined the theory of statistical learning and the experimental paradigm of the MMN to gain understanding of how cohesive sound patterns are transiently represented in brain signals.

We substantiated pattern recognition and violations through a modeling framework trained to learn transitional probabilities of the input sequences. The two kinds of wrong notes violated local (within-pattern) and global (between-pattern) melodic structures and were equally improbable within the overall sequence. We predicted that a wrong note (deviant) that violates the structure of learned melodic phrases (within-pattern deviant) would elicit a stronger brain response than a wrong note that does not violate the structure of any of the learned melodic phrases (between-pattern deviant). Thus, evidence for implicit learning of the statistical structure of the sequence and instantiation of the melodic patterns in working memory would be expected by an amplitude difference

of the MMN component. A larger amplitude MMN is expected for violations of the melodic sound patterns because MMN is based on the standard repeating regularity (Sussman, 2007).

This prediction is quite different from an alternative hypothesis, more common to MMN paradigms, which is that the probability of the deviant determines the amplitude of its MMN, irrespective of the transitional probability being violated (Sculthorpe & Campbell, 2011). If this hypothesis is correct, it predicts that there will be no difference in MMN amplitude between the within-pattern and between-pattern deviants because the probability of both types of deviants is equal. Thus, the MMN will distinguish factors that influence how sequential sound information is stored in memory.

METHODS

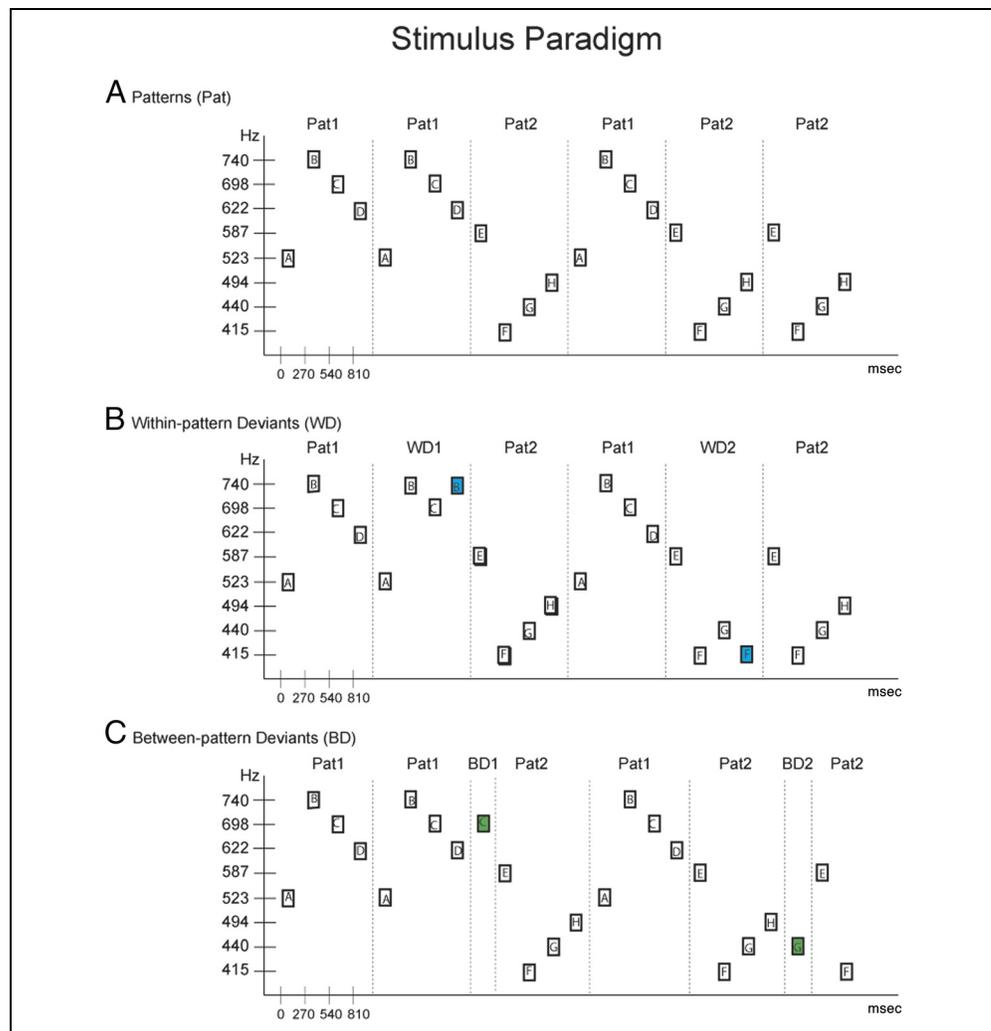
Participants

Ten normal hearing adults (five women, eight right-handed) ranging in age between 23 and 34 years ($M = 28$, $SD = 3$) participated in the study. Participants passed a hearing screening (20 dB HL or better at 500, 1000, 2000, and 4000 Hz) bilaterally and had no reported history of neurological disorders. The internal review board of the Albert Einstein College of Medicine, where the study was conducted, approved the procedures. All participants gave written informed consent before participating and were paid for their participation.

Stimuli

Eight pure tones were created in MATLAB (Mathworks, Inc., Natick, MA). Tones were 250 msec in duration (5 msec rise/fall time). Tones were calibrated to 58 dB(A) (2209 Sound Level Meter, Brüel & Kjær, Nærum, Denmark). Tones were the frequency equivalent of a musical note within the western musical system, occurring within a single octave ($G\#_4$: 415.30, A_4 : 440.00, B_4 : 493.89, C_5 : 523.25, D_5 : 587.33, $D\#_5$: 622.25, F_5 : 698.46, and $F\#_5$: 739.99 Hz). We hereforth denote tones using the letters A through H, but this representation is not related to the musical pitches. Tones were presented in two different four-tone patterns (Figure 1A). Pattern 1 (Pat1) was C_5 - $F\#_5$ - F_5 - $D\#_5$ (ABCD) and Pattern 2 (Pat2) was D_5 - $G\#_4$ - A_4 - B_4 (EFGH). A 20-msec silence occurred from the offset to onset of each tone. We used unfamiliar tone patterns so that listeners could not rely on prior knowledge of songs or musical advertisements. The two patterns each occurred with 50% probability in the sequence and were randomly concatenated (Figure 1A) with the constraint that consecutive deviant patterns had to be separated by at least two standard patterns. Tones were presented isochronously. Thus, there was no demarcation when one pattern terminated and the next one began. Tone sequences began with presentation of at least 15 standard patterns before any deviants occurred.

Figure 1. Schematic diagram of the stimulus paradigm. (A) Sequential tone patterns (Pat) are depicted with rectangles, labeled ABCD (Pat1) and EFGH (Pat 2). Frequency values of the tones are displayed in Hz on the ordinate, with timing between sounds (in milliseconds) shown on the abscissa. Patterns are demarcated by a dashed vertical line for display purposes only. (B) Within-pattern deviants (WD) are displayed for Pat1 and Pat2. WD1 (blue-filled rectangle) violates expected Pat1; WD2 (green-filled rectangle) violates expected Pat2. (C) Between-pattern deviants (BD) are displayed, following Pat1 and Pat2. Between-deviant violations occur outside the borders of the four-tone patterns, serving as global deviants to the overall sequence. Dashed lines demarcate the pattern borders for display purposes only.



The transitional probability (excluding the deviant tones) for tones within each pattern was 1.0 (e.g., A → B, B → C, C → D), and the transitional probability between patterns (switching from one pattern to the next) was .5 (e.g., D → A or D → E). The high probability of transitioning from one tone to the next within each pattern and the relatively low probability of transitioning from one pattern to the next permit the segmentation of the patterns into two distinct, statistically cohesive units.

Every so often, the regularity of the sequence was violated by one of two kinds of “wrong” notes (or “deviants”). Within-pattern deviants (WD1 and WD2) were the occurrence of the wrong tone replacing the fourth tone of the pattern (Figure 1B). Within-pattern deviants thus violated the expected structure of the melodic units (e.g., ABC**B** occurred when ABCD was expected; Figure 1B). Between-pattern deviants (BD1 and BD2) were the occurrence of the wrong tone after the conclusion of one of the patterns and before the onset of the next (Figure 1C). Between-pattern deviants thus did not alter the structure of the melodic units (e.g., ABCD**C** occurred when ABCDA or ABCDE was expected; Figure 1C).

Deviants occurred with a 1:5 probability ratio relative to the standards, meaning that deviant patterns comprised 16.66% of the total number of patterns in the sequence. Both types of deviants occurred equiprobably within the stimulus blocks such that each of the four deviant pattern types occurred approximately 4% of the time. However, the presence of the deviants in the sequence modified the overall transitional probabilities between tones (Figure 2).

There were four criteria for wrong notes. One of the criteria was that wrong notes were not new tones within the sequences. For the within-pattern deviants, the wrong note was the second tone of the pattern occurring in place of the fourth tone (W1: C₅-F₅-F₅-F₅; W2: D₅-G₄-A₄-G₄), and for the between-pattern deviants, the wrong note was the third tone of the preceding pattern (BD1: C₅-F₅-F₅-D₅-F₅; BD2: D₅-G₄-A₄-B₄-A₄). Thus, wrong notes were tones already used within the sequence and therefore would only be detected if the order of the tones in melodic units were maintained in working memory. A second criterion was that there was not a set trajectory for the deviants. That is, we used two sets of patterns

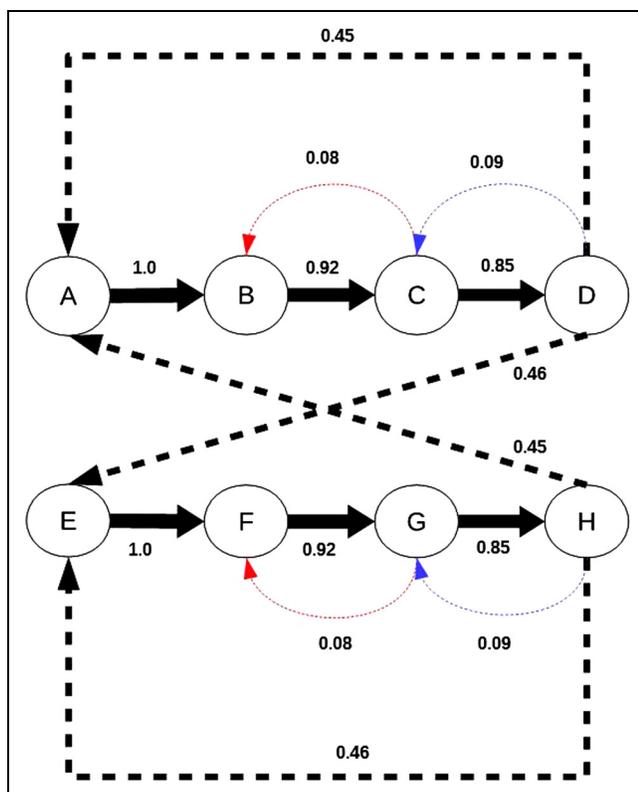


Figure 2. Schematic of the Markov model trained on the stimulus sequence shown in Figure 1. Circles represent states (tones), and arrows represent transitions, labeled with their first-order transition probabilities. Arrow thickness depicts strength of the transition probabilities. Transitions with probability exceeding a threshold of .83 are shown with thick solid arrows. Lower probability transitions are shown with thick broken arrows. Thin broken lines correspond to within- and between-pattern deviants. The slight differences in probability between the within- and between-pattern deviants are because the patterns containing between-pattern deviants also contribute to the probability of transitioning from the third tone of the pattern to the final tone. (Note: Some transitions were omitted due to space constraints.)

and deviants so that one deviant violated the pattern by ascending instead of descending to the fourth tone and the other by descending instead of ascending to the fourth tone. In this way, the direction to the wrong note could not, on its own, be used to cue a deviation. The third criterion was that the frequency distance between the expected note and the wrong note was three semitones for all deviants, and the final criterion was that the intervallic violation of the between-pattern deviants was as large as the intervallic violation of the within-pattern deviants. A backtracking constraint satisfaction algorithm was used to select the standard patterns and deviants according to the above criteria.

Procedures

Participants were seated in a comfortable chair in an electrically shielded, sound attenuated booth (IAC, Bronx, NY). Sounds were presented bilaterally through insert earphones (E-A-RTONE 3A, Indianapolis, IN). Participants

were instructed to ignore the sounds and watch a closed-captioned silent movie of their choice. Stimuli were presented in six blocks of stimuli. In total, 1140 of each standard pattern type and 114 of each deviant pattern type were presented. The presentation order of the six blocks was randomized across participants. Participants took a break at the midpoint to be disconnected from the EEG recording system and have a snack. The total session time was less than 2 hr, including electrode cap placement, instructions, and breaks.

EEG Recording and Data Reduction

EEG was recorded using a 32-channel electrode cap (Electro-Cap International, Inc., Eaton, OH), in the International modified 10–20 system, including electrodes placed on the left mastoid (LM) and right mastoid. The tip of the nose was used for the reference electrode and P09 for the ground electrode. A bipolar configuration was used to monitor the horizontal EOG between electrodes F7 and F8 and to monitor the vertical EOG with electrodes between FP1 and an external electrode placed below the left eye. Impedances were maintained below 5 k Ω . EEG and EOG were digitized at a sampling rate of 500 Hz with a bandpass of 0.05–100 Hz. The EEG was then filtered offline with a high-pass cutoff of 0.1 Hz and a low-pass cutoff of 30 Hz using a finite impulse response filter with zero phase shift and a roll-off slope of 24 dB/octave. Filtered EEG was segmented into 1500 msec epochs, including a 200-msec prestimulus period. The response to the first tone of the standard pattern was used to create the “standard” waveform, and the response to the first tone of the deviant pattern was used to create the “deviant” waveform. Thus, the epochs included all of the tones in the patterned units. Epochs were baseline-corrected before artifact rejection was applied with a criterion set at ± 75 V on all electrodes (EOG and EEG). On average, 15% of the overall epochs were rejected due to artifact. The remaining EEG epochs were averaged separately by stimulus type and then baseline-corrected to the prestimulus period.

The MMN was delineated in the grand mean difference waveforms (deviant minus standard). The peak of the MMN was first visually identified in the grand mean difference waveform at the Fz electrode where the signal-to-noise ratio is greatest. To statistically evaluate the presence of the MMN component, the data were re-referenced to the LM. A 50-msec window centered on the peak latency of the MMN was then used to obtain the mean amplitudes for the standard and deviant ERPs, in each individual, separately for each pattern and deviant type (944 msec was the peak latency for WD and 1216 msec peak latency for BD, as calculated from the first stimulus of the pattern). The within-pattern deviant coincided with the final tone of the pattern, and the between-pattern deviants occurred 270 msec after the end of a standard pattern (1216 – 944 = 272 msec).

A three-way omnibus repeated-measures ANOVA was calculated on the difference waveforms (deviant minus standard) to statistically verify the presence and scalp distribution of the MMN component (factors of Deviant type [within/between], Pattern type [1/2], and Electrode [Fz, Cz, Pz]). Where data violated the assumption of sphericity, degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity, and corrected degrees of freedom and p values were reported. For post hoc analyses, when the omnibus ANOVA was significant, Tukey's honestly significant difference for repeated measures was conducted on pairwise contrasts. Contrasts were reported as significantly different at $p < .05$. All statistical analyses were performed using Statistica 12 software (Statsoft, Inc., Tulsa, OK).

Modeling Methods

The modeling framework contains two stages. The first stage of the model is an algorithm that generates a probabilistic model of a stimulus sequence that enables the formation of strongly predictable, cohesive patterns by learning transition probabilities. The second stage of the model matches incoming stimuli to the patterns that have been learned through the statistical learning process and detects violations of pattern regularities. To represent the probabilistic relationships between stimuli in a temporal sequence, we used a Markov chain model to learn transition probabilities.

Chains of strongly connected (high transition probability) states in the Markov chain are analogous to patterns or melodies stored in a memory trace “dictionary” in the brain. Because we are using a first-order model, the memory trace would consist of pairs of strongly connected stimuli ($A \rightarrow B$, $B \rightarrow C$, and so on [the arrow indicates “is followed by”]).

Once the regularities in the sequence are learned, the next part of the process is to match incoming stimuli to the regularities learned to see if there is a violation of a strong prediction and define a transition probability threshold, b , which determines whether a particular transition expectation creates a strong predictive rule to which an MMN will be generated. For our model, we used a threshold of 0.825, which corresponds to the standard–deviant ratio in a typical MMN experiment. Incoming stimuli are compared with the expectations represented by the transition probabilities that are learned. If consecutive stimuli match the expectations of the statistical learning (e.g., $A \rightarrow B$), no MMN will be elicited. If the transition from the previous stimulus to the current stimulus does not match the prediction represented by the statistical learning, however, a large MMN is generated if and only if there is at least one transition in the graph from the previous stimulus that is greater than the threshold b . In other words, we predicted that an MMN is generated in response to a stimulus if and only if a strong expectation created by the previous stimulus is violated.

This is in line with the theory that MMN is a response to the violation of a specific memory trace, not just a global irregularity.

RESULTS

Prediction and Modeling Framework

Our main prediction was that the deviants occurring within the statistically cohesive patterns (within-pattern deviants) would elicit a strong MMN because the transition probabilities for the tones within the unit would be large, whereas the deviants occurring between patterns (between-pattern deviants) would elicit a less robust MMN because the deviants would violate a weaker probabilistic expectation (Figure 2). To formalize this experimental prediction, we designed a simple two-stage model to describe how the brain learns the statistical structure of sensory input and responds to violations of expectations.

The first stage of the model consisted of a first-order Markov model that learned transition probabilities from the same tone sequence heard by our participants in the experimental paradigm (Figure 2). The second stage of the model matched incoming stimuli to the regularities that have been learned through the statistical learning process and produced a binary surprise response to represent whether a strong expectation was violated. This process resulted in strong expectations for cohesive patterns (ABCD and EFGH) due to the high transition probabilities.

Although both within- and between-pattern deviants were equally unlikely, the model only responded to the violation of a strong expectation, which was not present for the between-pattern deviants. One of the important implications of this model result is that deviancy should be treated differently by the brain when violations are based on strong information units (e.g., within-pattern deviants), and this distinction would be apparent in the strength of the MMN response.

Experimental Results

We tested the predictions of the above statistical learning-based model. Table 1 summarizes the mean amplitudes of the standard and deviant ERP responses to all pattern types in the latency of the MMN component. Figure 3 displays the ERPs evoked by the standard and deviant waveforms for both deviant types. Figure 4 displays the difference waveform potentials (standard ERP subtracted from the deviant ERP) at all 32 electrodes and the global field power (Lehmann & Skrandies, 1980). Because there was no effect due to Pattern type, $F(1, 9) < 1$, $p = .64$, data were collapsed across pattern types for display only (Figures 3–4).

Within-pattern deviants elicited a significantly larger MMN than between-pattern deviants (main effect of Deviant type, $F(1, 9) = 9.9$, $p = .012$, $\eta_p^2 = .52$; Figures 3–4). Overall,

Table 1. ERP Mean Amplitudes (in μV) of the MMNs for the Deviant (Dev) and Standard (Std) Stimuli

Electrode	Deviant Type							
	Within-pattern 1		Within-pattern 2		Between-pattern 1		Between-pattern 2	
	Dev	Std	Dev	Std	Dev	Std	Dev	Std
Fz	-0.99 (1.97)	0.64 (0.72)	-1.86 (1.1)	1.13 (0.57)	-0.52 (2.31)	-0.23 (0.84)	-0.44 (1.17)	0.53 (0.79)
Cz	-0.55 (2.27)	0.64 (0.81)	-1.45 (1.4)	0.73 (0.53)	-0.50 (1.9)	-0.16 (0.96)	-0.09 (2.07)	0.37 (0.75)
Pz	0.85 (1.81)	-0.06 (0.82)	-1.13 (2.0)	0.21 (0.63)	0.21 (2.3)	-0.24 (1.04)	0.16 (2.1)	0.26 (0.79)
LM	2.16 (1.59)	-0.55 (0.45)	0.42 (1.4)	-0.19 (0.61)	1.17 (1.58)	0.11 (0.91)	1.3 (1.07)	0.68 (0.62)

SD in parentheses.

the scalp distribution of the response was consistent with a frontocentral maximum for MMN (main effect of Electrode, $F(1.52, 13.64) = 9.9$, $\epsilon = 0.75$, $p = .012$, $\eta_p^2 = .70$), with post hoc calculations showing a mean amplitude at Pz significantly smaller (less negative) than both Fz and Cz ($p < .001$) and no amplitude difference between Fz and Cz ($p = .18$). There was an interaction of Deviant type with Electrode, $F(1.27, 11.42) = 4.83$, $\epsilon = 0.75$, $p = .043$, $\eta_p^2 = .35$, due to the within-pattern deviants showing the fronto-

central distribution (Pz smaller than Fz and Cz), but between-pattern deviants having no electrode effect, no significant mean amplitude difference among Fz/Cz/Pz. The interaction of Pattern type with Electrode did not reach significance ($p = .06$), and there was no interaction among all factors (Deviant type/Pattern type/Electrode, $p = .21$). Thus, MMNs were from both transitions, but the within-pattern deviants elicited a significantly more robust response than the between-pattern deviants.

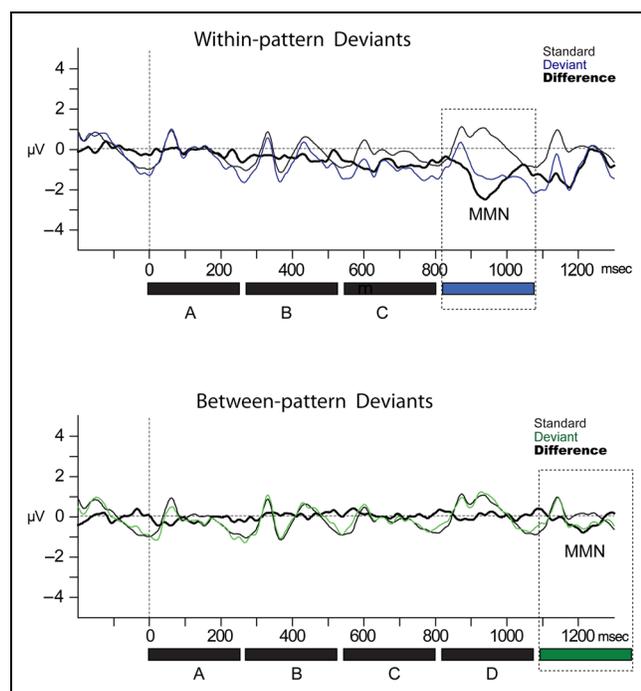


Figure 3. ERPs. ERPs evoked by the wrong notes occurring within melodic phrases (top row, thin blue line) and between melodic phrases (bottom row, thin green line) are displayed along with the ERPs evoked by the corresponding standards (thin black line). Color of the deviants is consistent with Figure 1 showing the stimulus paradigm. The subtraction (deviant – standard) waveforms are displayed by a thick black line. The MMN response is labeled and highlighted with a dashed black rectangle, corresponding with the timing of the deviant. The stimuli are depicted below the x-axis (shading is consistent with the traces) to show the timing of the responses to each of the tones within the patterns, along with the occurrence of the deviants.

DISCUSSION

The current results demonstrate implicit learning of stimulus statistics and transitional probability leading to detection and storage of musical phrases in auditory memory. Violations of the high (.85) transition probability (wrong notes occurring within the musical phrases) elicited a stronger electrophysiological response (larger MMN amplitude) than violations of the low (.45) transition probability (wrong notes occurring between musical phrases). The discrepancy between MMN amplitudes is unrelated to the probability of within-pattern versus between-pattern deviants, because each deviant occurred with equal probability within the tone sequence.

In most previous investigations on the effects of transitional probability, the probabilities of the deviant tones were manipulated. These previous studies showed violation of the prediction through the deviant: lower probability deviants evoked stronger responses than higher probability deviants (Koelsch et al., 2016; Paraskevopoulos et al., 2012; Furl et al., 2011). These studies were assessing the “surprising” nature of the deviants based on detection of the transitional probabilities in the sequence. That is, they were showing that the more unexpected the occurrence of a tone is (the lower probability of the deviant), the further away it is from what is predicted based on the stimulus statistics, results in a larger MMN response. Thus, in these previous studies violations occurred only for within-pattern deviants. Deviant responses represented a violation of the prediction based on the stimulus statistics. In the current study, in contrast, we did not manipulate the probability of occurrence for the

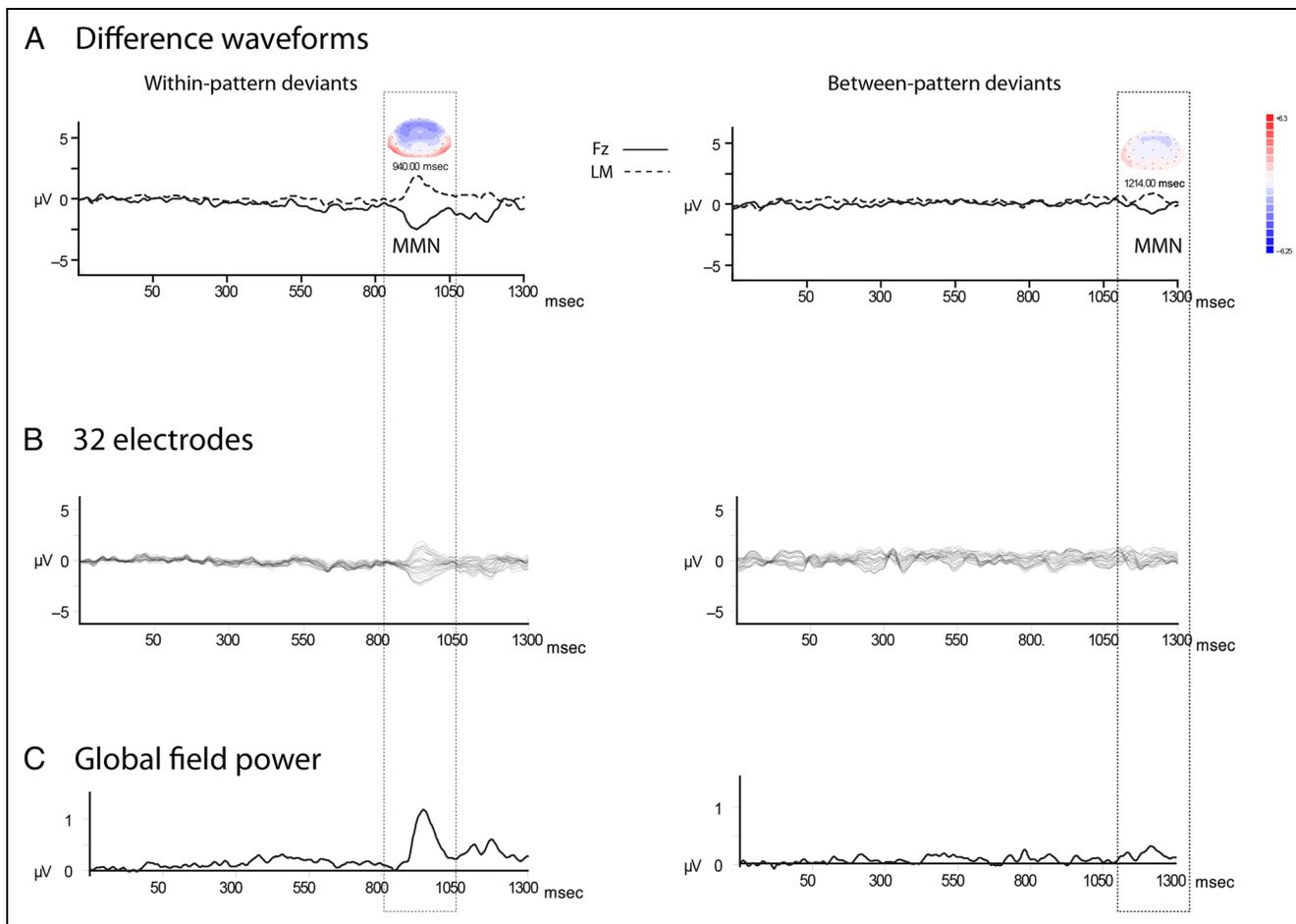


Figure 4. Difference waveforms. (A) Grand mean difference waveforms (deviant ERP minus standard ERP) are displayed showing the MMN component (labeled MMN), with the Fz (solid line) and LM (dashed line) electrodes overlain. The vertical dotted rectangle highlights the borders of the MMN. Patterns 1 and 2 are collapsed across standards and deviants for the within-pattern deviants (left column) and between-pattern deviants (right column). Peak latency of the components is indicated above the MMN, along with voltage maps showing the scalp distribution around the head (blue represents negative polarity, and red represents positive polarity). Waveforms were time-locked to the first tone of the patterns to show the timing of deviance detection for both pattern deviants. (B) Thirty-two electrodes. Data from all 32 electrodes overlain depict the scalp distribution across the length of the epoch and tones of the patterns. MMN is clearly present for the within-pattern deviants. (C) Global field power was calculated as the spatial root mean of the squared voltage deviations across electrode recordings from the entire scalp to provide a reference-free measure for component identification. The peak of power seen in each epoch (within and between deviants) corresponds to the MMN component depicted in the ERP waveforms (A and B).

within- and between-pattern deviances. This factor was held constant: The probability of occurrence for within- and between-pattern deviants was equal in the block. Instead, we used the transitional probabilities of the melodic tone patterns to assess how information is structured in memory. Thus, the strength of the response to the deviant was not based on the degree of surprise for the deviant but on the representation of the standard providing the basis for deviance detection.

Our results demonstrate that the unitized information (by melodic phrase) is stored in auditory memory because we show that MMN amplitude reflects a stronger response to a violation of the melodic tone patterns held in memory (within-pattern deviants) than to deviants that occur but do not alter the structure of the melodies (between-pattern deviants). This indicates that statistical learning may con-

tribute to the storing of temporal patterns as cohesive units in auditory memory, as predicted by the first-order Markov chain framework. These data thus provide new evidence of how sounds are organized to form informational units in memory that are useful for perception. However, not all statistical learning situations are first order, and some would require more complex models. For instance, in Symonds et al. (2017), there was a single standard pattern sequence that consisted of binary tones “XXXO.” The violations were longer and shorter sequences than the standard patterns (“XXXXO” and “XXO,” respectively). In this case, learning the tone pattern would require Markov models that go beyond first order.

MMN amplitude is thought to be affected by detectability of deviants (Gomes et al., 2000; Schroger, 1995; Tiitinen, May, Reinikainen, & Näätänen, 1994; Näätänen, 1988), and

thus, the amplitude may have been larger for the within-pattern deviants because they were easier to detect. It is possible that the between-pattern deviants may have been detected as being deviant by extending the length of the standard pattern from four tones to five tones. If this is the case, our results also suggest that changing an element of an established melodic pattern (within-pattern deviants) more strongly violates the expectation generated by the repeating tone patterns (standards) than extending the length of the deviant. This indicates that the melodic patterns were more strongly represented as repeating events in memory than the simple identity of the tones in the sequence.

The results are also interesting in light of the long-standing belief that “global” deviants elicit MMN. The conventional notion about MMN generation has been that any infrequent change occurring within a sequence will elicit MMN because it is different from the global sequence. However, more recent work has demonstrated that deviants are detected only when they violate a detected standard and not simply because a tone occurs infrequently in a stimulus block (Sussman et al., 2002, 2014; Sussman, 2007). The current results extend these results, indicating that MMN will be elicited when a specific, strong expectation is violated. Overall, our results demonstrate that informational units form the basis for processing the auditory scene and detecting novelty within it.

It is not immediately apparent how the brain would go about learning statistically cohesive temporal patterns from a biological standpoint. Xu, Jiang, Poo, and Dan (2012) provided evidence that suggests a possible mechanism for the storage of Markov chain-like memory traces. Rats were shown a visual stimulus that repeatedly moved from an initial starting point to a finishing point. During the stimulus presentation, neurons in primary visual cortex fired when the stimulus entered each of their respective receptive fields. After training on this motion sequence, only the initial stimulus (i.e., the starting point) was presented to the rat. Now, neurons from the receptive fields along the path fired in the original temporal order that they had been trained on and not only those generated by the initial stimulus, demonstrating that the rat had learned the temporal sequence.

The authors proposed the simplest kind of Hebbian learning, spike-timing-dependent plasticity, as the mechanism of this learning (Xu et al., 2012). When the neurons repeatedly fired in a particular order due to stimulus presentation, the connection between those neurons was strengthened in the order that they were initially activated, effectively creating a memory trace. This phenomenon is precisely what is described in our Markov chain model for MMN, wherein repeated sequences with high transition probabilities create a cohesive temporal memory trace that the brain can then use to make predictions and produce violation responses when the predictions are not met. Spike-timing-dependent plasticity is a domain-general feature of neurons and

neural circuits in many parts of the brain. It would thus be unsurprising to find statistical learning responses for various types of regularities, such as timbre and intensity in the auditory modality, as well as in multiple sensory modalities.

Acknowledgments

We thank Wei Wei Lee and Sufen Chen for their assistance with data collection and analysis. This research was funded by the National Institutes of Health (DC004263, E. S. S.).

Reprint requests should be sent to Elyse S. Sussman, Department of Neuroscience, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, or via e-mail: elyse.sussman@einstein.yu.edu.

REFERENCES

- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*, 170–176.
- Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*, 437.
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*, 133–150.
- Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*, 287–295.
- Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *Neuroimage*, *54*, 2267–2277.
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*, 453–463.
- Gervain, J., & Werker, J. F. (2013). Learning non-adjacent regularities at age 0;7. *Journal of Child Language*, *40*, 860–872.
- Gomes, H., Molholm, S., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G. J. (2000). Mismatch negativity in children and adults, and effects of an attended task. *Psychophysiology*, *37*, 807–816.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.
- Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports*, *6*, 19741.
- Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, *48*, 609–621.
- Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*, 91–121.
- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494.
- Mill, R., Coath, M., Wennekers, T., & Denham, S. L. (2011). A neurocomputational model of stimulus-specific adaptation

- to oddball and Markov sequences. *PLoS Computational Biology*, 7, e1002117.
- Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102, 156–160.
- Myers, J., Jusczyk, P. W., Kemler Nelson, D. G., Charles-Luce, J., Woodward, A. L., & Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language*, 23, 1–30.
- Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology*, 26, 117–163.
- Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory-cognitive core of cognitive processes. *Brain Research Reviews*, 64, 123–136.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, 24, 283–288.
- Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia*, 50, 341–349.
- Saarinen, J., Paavilainen, P., Schoger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, 3, 1149–1151.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Schroger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, 32, 55–65.
- Schroger, E., Näätänen, R., & Paavilainen, P. (1992). Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain. *Neuroscience Letters*, 146, 183–186.
- Sculthorpe, L. D., & Campbell, K. B. (2011). Evidence that the mismatch negativity to pattern violations does not vary with deviant probability. *Clinical Neurophysiology*, 122, 2236–2245.
- Sussman, E. (2007). A new view on the MMN and attention debate: The role of context in processing auditory events. *Journal of Psychophysiology*, 21, 164–175.
- Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, 36, 22–34.
- Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Brain Research. Cognitive Brain Research*, 13, 393–405.
- Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The five myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography*, 27, 553–564.
- Symonds, R., Lee, W., Kohn, A., Schwartz, O., Witkowski, S., & Sussman, E. S. (2017). Distinguishing stimulus specific adaptation and predictive coding hypotheses in auditory change detection. *Brain Topography*, 30, 136–148.
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10, 21.
- Teinonen, T., & Huotilainen, M. (2012). Implicit segmentation of a stream of syllables based on transitional probabilities: An MEG study. *Journal of Psycholinguistic Research*, 41, 71–82.
- Tervaniemi, M., Rytönen, M., Schroger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 8, 295–300.
- Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, 372, 90–92.
- van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Näätänen, R., & Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: An event-related brain potential study. *Journal of Cognitive Neuroscience*, 18, 1292–1303.
- van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of sequential sounds—An event-related potential study comparing musicians and nonmusicians. *Journal of Cognitive Neuroscience*, 16, 331–338.
- Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, 7, 270–276.
- Xu, S., Jiang, W., Poo, M., & Dan, Y. (2012). Activity recall in a visual cortical ensemble. *Nature Neuroscience*, 15, 449–455.