Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 5, 1119–1130

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.5.1119

Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences

Sarah C. Creel, Elissa L. Newport, and Richard N. Aslin
University of Rochester

Human listeners can keep track of statistical regularities among temporally adjacent elements in both speech and musical streams. However, for speech streams, when statistical regularities occur among nonadjacent elements, only certain types of patterns are acquired. Here, using musical tone sequences, the authors investigate nonadjacent learning. When the elements were all similar in pitch range and timbre, learners acquired moderate regularities among adjacent tones but did not acquire highly consistent regularities among nonadjacent tones. However, when elements differed in pitch range or timbre, learners acquired statistical regularities among the similar, but temporally nonadjacent, elements. Finally, with a moderate grouping cue, both adjacent and nonadjacent statistics were learned, indicating that statistical learning is governed not only by temporal adjacency but also by Gestalt principles of similarity.

How do listeners organize and learn a patterned sequence of elements? Recent studies of a mechanism we have called *statistical learning* have shown that adults, young children, and infants are capable of computing transitional probabilities among adjacent syllables in rapidly presented streams of speech and of using these statistics to group syllables into word-like units (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997).[1] We have also shown that adults and infants can perform the same type of learning on streams of tones, showing an equivalent ability to use the statistical consistencies among adjacent tones to group them into small patterned melodies (Saffran, Johnson, Aslin, & Newport, 1999), as well as on sequences of patterned visual elements (Fiser & Aslin, 2002; see also Kirkham, Slemmer, & Johnson, 2002) and in patterned serial visual-motor responding (Hunt & Aslin, 2001).

An important question concerns the status of temporal adjacency in this and other types of pattern learning processes. In many domains, patterned relationships occur among elements that are not immediately adjacent to one another. Learning these domains therefore necessarily involves acquiring such nonadjacent relationships. For example, in some languages, words contain regular patterns among sounds that are not adjacent (e.g., in Hebrew and Arabic, verbs are built from a consonant pattern, such as *k-t-b,* with vowel patterns inserted between the consonants to indicate tense and number), and in music, the main melodic line may consist of notes that are not all adjacent (e.g., bass notes may intervene between notes of the melody). However, in research on sequential order, learners have shown quite limited ability to acquire patterns among nonadjacent elements, with a clear advantage for elements that are temporally adjacent, and under certain circumstances, learners have shown very restricted success in acquiring nonadjacent regularities (Cleeremans, 1993; Cleeremans & McClelland, 1991; Newport & Aslin, 2004).

One way of resolving this discrepancy is to consider the possibility that, although nonadjacent regularities may in general be difficult to acquire, certain types of nonadjacent patterns may be easier to learn than others, and those that are difficult to learn tend not to appear in language or music. Indeed, although languages exhibit some nonadjacent relationships, a central finding of modern linguistics has been that such nonadjacent relations are quite selective and limited in languages of the world. A main enterprise of theoretical linguistics of all flavors has been to capture these limitations in a set of principles or universal constraints (Chomsky, 1965, 1981, 1995; Gazdar, Klein, Pullum, & Sag, 1985; Pollard & Sag, 1994). In the present work, we asked whether at least some of these constraints might be broader than language and might arise from more general constraints on the learning of sequential patterns.

In recent work, Newport and Aslin (2004) reported a potentially informative set of findings regarding the types of nonadjacent patterns that learners can and cannot readily acquire. In streams of synthetic speech, they found that human listeners were unable to extract consistencies among syllables that were nonadjacent (with one unpredictable syllable intervening). However, when statistical regularities occurred among nonadjacent phonemic segments—

Sarah C. Creel, Elissa L. Newport, and Richard N. Aslin, Department of Brain and Cognitive Sciences, University of Rochester.

This research was supported in part by National Science Foundation (NSF) Grant SBR-9873477 to Richard N. Aslin and Elissa L. Newport, National Institute of Health (NIH) Grant DC00167 to Elissa L. Newport, NIH Grant HD37082 to Richard N. Aslin, and NSF Graduate Research Fellowship to Sarah C. Creel.

Correspondence concerning this article should be addressed to Sarah C. Creel, Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, NY 14627. E-mail: screel@bcs.rochester.edu

[1] More technically, we have shown that learners compute a conditionalized statistic that tracks the consistency with which elements occur together and in a particular order, baselined against individual element frequency. *Transitional probability* is a particular type of temporally ordered conditional probability, first used for psycholinguistic materials by Miller and Selfridge (1950), but our findings are also compatible with the claim that learners might be computing another closely related statistic, such as mutual information or conditional entropy.

either among consonants (skipping over the intervening vowels) or among vowels (skipping over the consonants)—listeners could acquire the regularities quite readily (Newport & Aslin, 2004). This contrast in the learning of nonadjacent elements is consistent with a similar contrast in the structure of natural languages: Linguistic rules involving nonadjacent syllables are quite uncommon in languages of the world, whereas those involving nonadjacent segments are quite common (see Newport & Aslin, 2004, for discussion).[2]

Newport and Aslin (2004) suggested two hypotheses to explain this syllable–segment contrast. One possibility is a constraint that is specific to speech: Human listeners (adults in this particular study) might only compute statistical regularities among segments and not among syllables; syllable regularities might not be acquired at all or might be constructed indirectly from the segment regularities. On this interpretation, the selectivities in learning are not tied to nonadjacency, but they are specific to speech, which (unlike many other types of temporally sequenced stimuli) is comprised of a hierarchically organized set of elements that may be processed and remembered in distinct ways.

A second possible explanation, however, is that a more general perceptual constraint might be responsible for these selectivities of learning. Newport and Aslin (2004) suggested that certain Gestalt principles of perception, if applied to long-term learning, might help to explain these differences in learning nonadjacencies. Their findings suggest that learning relationships between adjacent elements is relatively simple, whereas learning relationships between nonadjacent elements is relatively difficult. Why, then, are nonadjacent segment regularities learned so easily? In all the nonadjacent segment languages Newport and Aslin studied, the nonadjacent segments that were regularly related to one another were of one element type, whereas the intervening segments were of another element type. In one case, the language was composed of patterns among consonants within a word, while the intervening vowels varied. In another case, the language was composed of patterns among the vowels, while the intervening consonants varied. An intuitive way of thinking about the ease of learning such regularities is that, when patterned elements are of one type and the unrelated intervening elements are of another type, noticing and storing the regularities among elements of like kind is much easier, and the usual difficulty of nonadjacency is ameliorated.

This effect is reminiscent of the Gestalt principle of similarity, although that principle was formulated to explain grouping in perception, not learning. According to the Gestalt principle of similarity (Wertheimer, 1923/1938, 1944), listeners tend to perceive elements that are physically similar to one another as grouped together and more closely related than their objective temporal or spatial distances would suggest (see also Bregman, 1990, on auditory streaming). Our results with temporally interleaved consonants and vowels suggest that Gestalt principles of similarity may also constrain the statistical learning of speech. In particular, principles of element similarity might interact with temporal adjacency in determining what types of patterned regularities can be acquired.

In the present series of experiments, we pursued these questions by studying the statistical learning of nonspeech stimuli. Our stimuli are patterned sequences of tones, similar to those used in the study by Saffran, Johnson, Aslin, and Newport (1999). However, in the present experiments, the patterned statistical regularities are among tones that are not temporally adjacent. In the first experiment, we asked whether these nonadjacent regularities are easily learned (in contrast to our findings with speech stimuli) or rather whether (as in speech) these nonadjacent regularities are very difficult to acquire. In subsequent experiments, we introduced a grouping cue (in Experiment 2, the grouping cue is pitch; in Experiments 3 and 4, the grouping cue is timbre), asking whether nonadjacent regularities become easier to learn when they are represented in stimuli of one type (e.g., high pitch), while the intervening stimuli are of another type (e.g., low pitch). Such contrasts have been studied in a variety of perceptual tasks, but little is known about their effects on implicit learning. In our conclusions, we return to the question of the similarities and differences between statistical learning of speech and tones and what these findings might mean for understanding pattern learning in general and language acquisition in particular.

Experiment 1

Previous statistical learning studies have shown that, for speech, certain types of nonadjacent regularities can be readily learned (e.g., among the consonants of consonant-vowel syllables), whereas others (every second syllable) are difficult to learn (Newport & Aslin, 2004). For streams of tones, Saffran et al. (1999) found that high transitional probabilities between adjacent tones served as an effective source of information for grouping, and performance was the same as for streams of speech syllables. However, it is unknown whether regularities between nonadjacent elements of tone sequences can be learned. We explored whether nonadjacent tone relationships can be learned using a series of tone triplets. Two sets of tone triplets were constructed and then temporally interleaved as shown in Figure 1; thus, the first tone of the first triplet was followed by the first tone of the second triplet, then the second tone of the first, the second of the second, and so on.

Table 1 presents the structure of the materials in pitch class notation. The items in bold in the section for Experiment 1 shows the two sets of tone triplets used in this experiment. The two sets of triplets were all in the same frequency range, C4 to B4. As in previous experiments on statistical learning with both linguistic and nonlinguistic stimuli, these triplets occurred more often and more consistently during exposure than did other three-element sequences embedded in the stream of tones. There were four triplets, divided into two sets of two triplets each. Within a triplet, the first tone was always followed by the second tone, and the second by the third, creating transitional probabilities of 1.0 across the two tone pairs within a triplet. At the end of a triplet, either of two triplets of this set could follow, creating a transitional probability of .5 from the third tone to the first tone of the next triplet. However, the tones of each triplet were not temporally contiguous but rather were interleaved with the tones of the other set of

[2] Natural languages do not commonly construct words out of a nonadjacent syllable pattern; most phonological patterns occur among adjacent phonetic segments or syllables. In some languages, such as Tagalog, words may consist of two syllables with inflections (called *infixes*) that can be inserted between the two syllables. However, the two syllables also often occur adjacent to one another, and evidence from child language acquisition has suggested that they are acquired first in their adjacent form (Slobin, 1973).

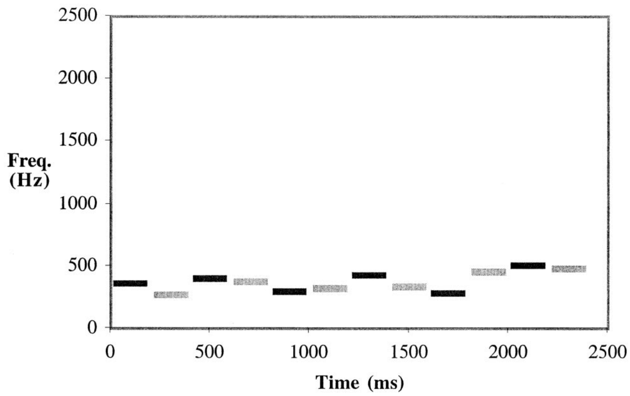

Figure 1. Frequency × Time graph of Experiment 1 stimuli. Each bar depicts the frequency (freq.) and duration of a tone. Odd-numbered tones are indicated by black bars; even-numbered tones are indicated by gray bars.

triplets. Thus, transitional probabilities of temporally adjacent tones were .5 both within and between triplets. In short, the only grouping by transitional probabilities was among the nonadjacent tones of a single triplet. All other transitional probabilities were .5.

To determine whether participants had learned these nonadjacent triplet groupings after exposure, they were asked to specify which of two alternatives sounded more familiar: a correct triplet or a misordered triplet. If participants were able to distinguish reliably correctly ordered from incorrectly ordered triplets, then they must have acquired sensitivity to the transitional probabilities in the sequences comprised of nonadjacent tones. We also included items to test participants' knowledge of the less predictive transitional probabilities (.5) between temporally adjacent tones compared with misordered tones. By measuring participants' knowledge of highly predictive nonadjacent tone sequences and less predictive adjacent tone sequences, we determined the relative importance of adjacency–nonadjacency and variations in predictiveness to the statistical learning of tone sequences.

Method

Participants

Twelve students at the University of Rochester participated in this experiment. To avoid the possibility that music experience might bias performance in the task, participants were selected who had not taken part in music lessons, music classes, or music performance groups since the 10th grade and who all identified themselves as nonmusicians (as in Saffran et al., 1999). No participants reported a history of hearing difficulties. All participants received $7.50 for participation.

Table 1
Stimuli Used in Experiments 1–4

Experiment	Test target items	Test foils
Experiment 1		
Nonadjacent	$\mathbf{F_4\_G_4\_D_4\_}$	$*C\#_4\_G\#_4\_B_4\_$
	$\mathbf{G\#_4\_C\#_4\_B_4\_}$	$*D_4\_G_4\_F_4\_$
	$\mathbf{\_C_4\_F\#_4\_D\#_4}$	$*\_A_4\_E_4\_A\#_4$
	$\mathbf{\_E_4\_A_4\_A\#_4}$	$*\_D\#_4\_F\#_4\_C_4$
Adjacent	$D_4D^{\#}_4F_4F^{\#}_4G_4C_4$	$*B_4A_4G^{\#}_4A^{\#}_4C^{\#}_4E_4$
	$D_4A^{\#}_4F_4E_4G_4A_4$	$*B_4F^{\#}_4G^{\#}_4D^{\#}_4C^{\#}_4C_4$
	$B_4D^{\#}_4G^{\#}_4F^{\#}_4C^{\#}_4C_4$	$*D_4A_4F_4A^{\#}_4G_4E_4$
	$B_4A^{\#}_4G^{\#}_4E_4C^{\#}_4A_4$	$*D_4F^{\#}_4F_4D^{\#}_4G_4C_4$
Experiment 2		
Nonadjacent	$\mathbf{F_6\_G_6\_D_6\_}$	$*C^{\#}_6\_G^{\#}_6\_B_6\_$
	$\mathbf{G^{\#}_6\_C^{\#}_6\_B_6\_}$	$*D_6\_G_6\_F_6\_$
	$\mathbf{\_C_4\_F^{\#}_4\_D^{\#}_4}$	$*\_A_4\_E_4\_A^{\#}_4$
	$\mathbf{\_E_4\_A_4\_A^{\#}_4}$	$*\_D^{\#}_4\_F^{\#}_4\_C_4$
Adjacent	$D_6D^{\#}_4F_6F^{\#}_4G_6C_4$	$*B_6A_4G^{\#}_6A^{\#}_4C^{\#}_6E_4$
	$D_6A^{\#}_4F_6E_4G_6A_4$	$*B_6F^{\#}_4G^{\#}_6D^{\#}_4C^{\#}_6C_4$
	$B_6D^{\#}_4G^{\#}_6F^{\#}_4C^{\#}_6C_4$	$*D_6A_4F_6A^{\#}_4G_6E_4$
	$B_6A^{\#}_4G^{\#}_6E_4C^{\#}_6A_4$	$*D_6F^{\#}_4F_6D^{\#}_4G_6C_4$
Experiments 3–4		
Nonadjacent	$\mathbf{F_{T1}\_G_{T1}\_D_{T1}\_}$	$*C^{\#}_{T1}\_G^{\#}_{T1}\_B_{T1}$
	$\mathbf{G^{\#}_{T1}\_C^{\#}_{T1}\_B_{T1}\_}$	$*D_{T1}\_G_{T1}\_F_{T1}$
	$\mathbf{\_C_{T2}\_F^{\#}_{T2}\_D^{\#}_{T2}}$	$*\_A_{T2}\_E_{T2}\_A^{\#}_{T2}$
	$\mathbf{\_E_{T2}\_A_{T2}\_A^{\#}_{T2}}$	$*\_D^{\#}_{T2}\_F^{\#}_{T2}\_C_{T2}$
Adjacent	$D_{T1}D^{\#}_{T2}F_{T1}F^{\#}_{T2}G_{T1}C_{T2}$	$*B_{T1}A_{T2}G^{\#}_{T1}A^{\#}_{T2}C^{\#}_{T1}E_{T2}$
	$D_{T1}A^{\#}_{T2}F_{T1}E_{T2}G_{T1}A_{T2}$	$*B_{T1}F^{\#}_{T2}G^{\#}_{T1}D^{\#}_{T2}C^{\#}_{T1}C_{T2}$
	$B_{T1}D^{\#}_{T2}G^{\#}_{T1}F^{\#}_{T2}C^{\#}_{T1}C_{T2}$	$*D_{T1}A_{T2}F_{T1}A^{\#}_{T2}G_{T1}E_{T2}$
	$B_{T1}A^{\#}_{T2}G^{\#}_{T1}E_{T2}C^{\#}_{T1}A_{T2}$	$*D_{T1}F^{\#}_{T2}F_{T1}D^{\#}_{T2}G_{T1}C_{T2}$

Note. Items in bold are the tone triplets. Letters are note names; numbers following letters denote octave of the note (e.g., middle C = C_4, B above middle C = B_4, and C two octaves above middle C = C_6). For Experiment 3, T1 is the timbre with two partials, and T2 is the timbre with three partials.

Stimuli

All sound stimuli were generated in SoundEdit 16.v2 and were recorded directly to Sony minidisc for presentation to individual participants. Each participant listened via Koss TD61 stereo headphones to an auditory sequence of familiarization materials (described below) for approximately 22 min and immediately afterward took a test on pairs of items from the familiarization materials.

Exposure. The exposure consisted of a series of tones comprised of constrained random orderings of four three-tone sequences (called *triplets*). Each triplet, listed in bold in the upper third of Table 1, was a unique set of three 200-ms sine tones. Each tone had onset and offset amplitude ramps of 10 ms to eliminate transient noises and followed the preceding tone without any intervening silent interval. Triplets were interleaved and randomly ordered as described below, with the constraint that the same triplet could only be repeated in immediate succession once before a different triplet intervened.

There were two sets of triplets. One set consisted of two triplets that occurred on odd-numbered tones, and the other set consisted of two triplets that occurred on even-numbered tones. The odd and even sets of tones were temporally interleaved; thus, a tone of one odd triplet was always preceded and followed by a tone of an even triplet. Hereafter, odd-numbered tones are denoted by capital letters and even-numbered tones by lower-case letters; thus, their combination is represented as OuPvQwRuSvTw . . . . (see Figure 2). No note was used in more than one triplet, and no octave-equivalent tones were used in different triplets.

The triplets were arranged in random order, with the constraint that each of the two triplets in a set followed either itself or the other triplet in that set an equal number of times and each triplet in that set was interleaved an equal number of times with each triplet in the other set. Thus, the nonadjacent transitional probabilities from one tone within a triplet to another tone within the same triplet, depicted in Figure 2, were always 1.0, and from the last tone in one triplet to the first tone in the next were always .5. The immediately adjacent transitional probabilities, also depicted in Figure 2, remained flat at .5 across the entire exposure, because a given tone was equally likely to be followed by one of two tones in the other set. Thus, although there were no consistent triplet patterns in the adjacent tones of the combined sequence, the combined sequence did contain some regularities in the "legal sequences" of tones that could occur (i.e., transitional probabilities = .5).

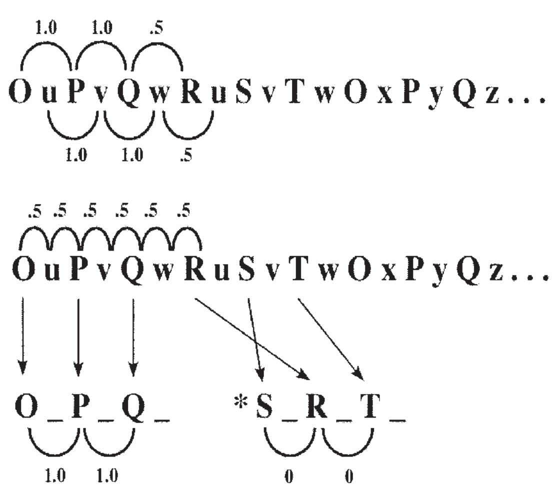

Figure 2. Experiment 1: Schematic of exposure materials and nonadjacent test items. The even set of tones is indicated in lowercase letters; the odd set of tones is indicated by capital letters. Transitional probabilities are indicated for nonadjacent tones (two letters away) and adjacent tones (one letter away). The asterisk indicates an incorrect ordering of the elements.

There were three blocks in the exposure stream, presented in sequence, which differed only in the particular randomization of the tone triplets. Each block lasted a total of 7 min and 13.2 s and contained 180 presentations of each of the triplets. Thus, the exposure totaled 22 min, during which each tone triplet was presented 540 times.

Test. Two types of two-alternative forced-choice (2AFC) test items were created and are listed in the upper third of Table 1. The first type of test item, the *nonadjacent-tone sequence,* was designed to assess learning of nonadjacent statistics within each set. These items consisted of a *nonadjacent triplet* (e.g., O_P_Q_ , where _ refers to the temporal interval of the intervening tone in the other set, with tone-to-tone transitional probabilities of 1.0), as contrasted with a *nonadjacent nontriplet,* created by inverting the order of two of the tones (e.g., *S_R_T_, whose within-set tone-to-tone transitional probabilities were 0). Each triplet or nontriplet contained 200 ms of silence (denoted in text by the underscore character) between tones, in the places where, during the exposure, a tone from the other set would have occurred. (Note that eliminating the silences would have resulted in a doubling of the presentation rate of the tones from exposure to test, possibly obscuring the identity of the triplet.) A triplet from one set (O_P_Q_) was always paired with a scrambled foil (*S_R_T_) created from the other triplet (R_S_T_) in the same set, to eliminate any possible effect of (odd or even) set preference. Correct and incorrect triplets had no pitches in common. Two triplet/nontriplet pairs per set × two sets × two orders (triplet first or nontriplet first) yielded eight unique test items, each of which was repeated once for a total of 16 nonadjacent test items.

The second type of test item was an *adjacent-tone sequence.* The adjacent statistics in the combined (i.e., interleaved) sequence were uniform (transitional probabilities = .5). Thus, there was no adjacent information for segmenting the sequence into groups of tones. However, if participants extracted the moderate transitional probabilities (.5) that were embedded in the sequence, then they should have been able to discriminate a familiar subset of this sequence from a novel sequence that had never occurred during the learning phase (transitional probabilities = 0). In these adjacent test items, a legal sequence of six interleaved tones from the two sets, such as QwOuPv (all adjacent transitional probabilities = .5), was contrasted with an illegal sequence, such as *TyRzSx (all adjacent transitional probabilities = 0). Note that neither of these six-tone test sequences are composed of two legal (occurring) triplets. Rather, they contain two part-triplets (Q–OP and w–uv or T–RS and yz–x). The rationale for using two interleaved part-triplets in the test items was to prevent participants from succeeding on the test by relying solely on the correctness or incorrectness of a triplet from a single set. For example, if we had contrasted two interleaved triplets (e.g., OuPvQw) with two randomly ordered triplets (e.g., SxRzTy), participants could have discriminated these two test items by attending to a within-set triplet (e.g., O_P_Q_ vs. *S_R_T_) rather than to adjacent tones in the test items. Use of part-triplets enabled both the legal and the illegal test items to contain within-set (nonadjacent) statistics that were matched (differing only in the ordering of 0.5 and 1.0 transitional probabilities). Thus, to differentiate correctly between these six-tone test items, listeners must have learned the difference in adjacent transitional probabilities (.5 in the interleaved part-triplets and 0 in the randomly interleaved part-triplets).

All sequences began with a tone from the odd set. Correct and incorrect sequences had no pitches in common. Four legal/illegal pairs × two orders yielded eight unique test items, which were each repeated once to generate a total of 16 adjacent-tone test items.

Two random orders of each section of the test (a block of 16 test trials) were generated, resulting in a 32-item test. Half of the participants received the nonadjacent (three-tone, within-set) items first, and the other half received the adjacent (six-tone) items first.

Procedure

In this and all following experiments, participants were instructed to “listen attentively to the sequence of tones,” and that a test would follow but were given no other instruction as to the structure of the exposure stimuli. They then were given a 32-item two-alternative, forced-choice test, consisting of 16 nonadjacent (3-tone) and 16 adjacent (6-tone) pairs. Participants were asked to select the item (first or second) that was “most like what they heard in the listening part of the experiment.” Responses were recorded using a printed test form. During both exposure and test, participants were allowed to adjust the volume of the stimuli to the desired level.

Results

Figure 3 presents the mean scores on the nonadjacent as compared with the adjacent test items. A 2 $\times$ 2 repeated-measures analysis of variance (ANOVA) was conducted on the data, with Test Item Type (adjacent or nonadjacent) as a within-participants factor and Test Order (adjacent first or nonadjacent first) as a between-participants factor.

There was a significant effect of Test Item Type, $F(1, 10) = 10.00$, $p = .01$. The average score ($M = 70.3\%$, $SE = 4.6\%$) on the adjacent tone test items was higher than the average score ($M = 52.6\%$, $SE = 3.6\%$) for the nonadjacent tone test items. There was no effect of Test Order, $F(1, 10) = .03$, $p = .8643$, and no significant interaction, $F(1, 10) = 2.22$, $p = .17$. Two-tailed t tests comparing the means for each test item type to chance (50% correct) revealed that participants exceeded chance only on the adjacent tone test items, $t(11) = 4.46$, $p = .001$. The nonadjacent tone test items did not exceed chance.

Discussion

These results suggest that listeners are not readily able to learn highly consistent relationships between nonadjacent tones. Despite the high predictiveness of the nonadjacent tone triplets, listeners could not distinguish these coherent tone orders with transitional probabilities of 1.0 from scrambled tone orders with transitional probabilities of 0. However, they were able to discern the partially predictive orderings of adjacent tones with transitional probabilities of .5 from scrambled tone orders with transitional probabilities of 0, even though there were no “dips” in these adjacent transitional probabilities as in previous statistical learning experiments (Saffran et al., 1996), and the part-triplet test items crossed the boundaries of the coherent, interleaved nonadjacent triplets. The results therefore appear to reflect an inability to learn nonadjacent statistics, mirroring results obtained by Cleeremans and McClelland (1991) in a serial reaction time paradigm and also mirroring results from human listeners presented with nonadjacent patterns of syllables (Newport & Aslin, 2004).

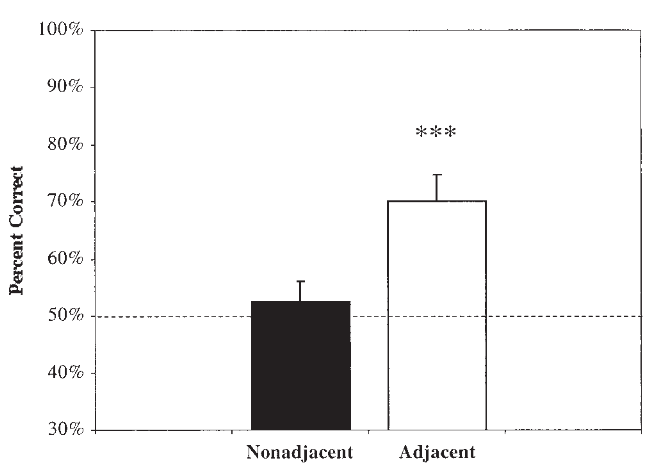

Figure 3. Experiment 1 test scores by item type. Error bars are standard errors. $^{***}p \leq .001$.

Experiment 2

Human listeners, although unable to readily acquire patterns among nonadjacent syllables, are adept at learning regular patterns among nonadjacent consonant or nonadjacent vowel segments. Newport and Aslin (2004) suggested two hypotheses to explain these results. First, this effect might be special to speech, arising from the tendency for learners to compute patterned relationships among certain types of speech elements (phonemic segments) but not among others (syllables). However, the results of the foregoing experiment suggest that this hypothesis may not be the correct account: Experiment 1 suggests that the difficulty of learning patterns among nonadjacent elements appears to extend to tones as well as to certain types of speech elements and therefore requires a more general account, not specific to speech alone.

Newport and Aslin’s (2004) second hypothesis was that the effects might be due to a more general perceptual constraint, involving Gestalt principles of similarity affecting the ease or difficulty of pattern learning. In particular, they suggested that learning nonadjacent relationships of any kind was relatively difficult, but this difficulty could be overcome if there was similarity among the nonadjacent elements (and difference from the intervening adjacent elements). In the speech experiments, this involved learning patterned relationships among elements of one type (e.g., consonants) while skipping over unrelated elements of another type (e.g., vowels). In Experiment 2, we pursued this hypothesis using our tone stimuli.[3]

The design of Experiment 2 was to present the same stimuli as Experiment 1, except that a grouping or similarity cue was introduced to differentiate two sets of nonadjacent stimuli that are patterned together within each set. Two literatures—one on Gestalt psychology and the other on auditory streaming—have suggested that such a manipulation might affect learning. Wertheimer (1923/1938, 1944) and other Gestaltists of the early 20th century outlined a number of grouping cues that would lead elements to be perceived as forming a group or coherent sequence. All other

[3] Our aim here was not to draw an analogy between tones and syllables or between tones and segments. Rather, given our hypothesis, we aimed at a more general similarity between our speech and tone materials: In the one case, the adjacent and nonadjacent elements were all of the same type, whereas in the other case, the interleaved nonadjacent elements were of two perceptually differentiated types. If we could achieve outcomes in the present series of studies that were comparable to these previous two speech cases, then we might have an explanation for both sets of effects—not in terms of differences between alternating syllables versus alternating segments but in terms of a constraint that enables nonadjacent learning only when the interleaved elements fall into two perceptually distinct categories.

things being equal, elements that occur immediately after one another are perceived as forming a group or melody (Wertheimer, 1944). However, temporal adjacency is not always paramount. Wertheimer (1944) noted that, in sequences of tones, if some of the tones are closer in frequency to one another than to other tones, they tend to group together. A more recent literature on auditory scene analysis, advanced by Bregman and others (e.g., Bregman, 1990; Bregman & Campbell, 1971; Bregman, Liao, & Levitan, 1990; Dannenbring & Bregman, 1976; Darwin, 1997), extends the study of Gestalt-style cues to the problem of assigning sounds to various environmental sources. Sounds that are spatially closer or spectrally more similar to one another tend to preferentially group together into auditory "streams," regardless of whether these sounds are temporally adjacent to one another.

However, both the Gestalt literature and the auditory scene literature focus on perceptual segregation; neither Gestalt principles nor streaming have been extensively investigated with respect to learning (but see Köhler, 1941; Rescorla & Furrow, 1977; and Rescorla & Gillan, 1980, for evidence that perceptual similarity may affect associative learning and Pavlovian conditioning).[4] In the present experiment, we added a perceptual cue to increase the similarity among the nonadjacent elements that are related to one another and to differentiate these elements from those that intervene. Specifically, we used pitch range—high or low—to mark this difference among the sets of tones. Our question was whether the nonadjacent patterns can then be learned. In Experiment 2, then, we used the stimuli from Experiment 1 but differentiated the octaves in which the two sets of triplets were presented. One set of triplets was kept in the same pitch range as in Experiment 1; the other set of triplets was produced two octaves higher. As before, the two sets of triplets were temporally interleaved; thus, high and low tones always alternated (see Figure 4). In all other ways— the statistical structure of the materials and of the test items, the duration of elements and items— Experiment 2 was identical to Experiment 1. The question of interest is whether the introduction of this pitch cue, making more similar the nonadjacent tones that are statistically patterned together, is adequate to produce learning of these nonadjacent patterns.

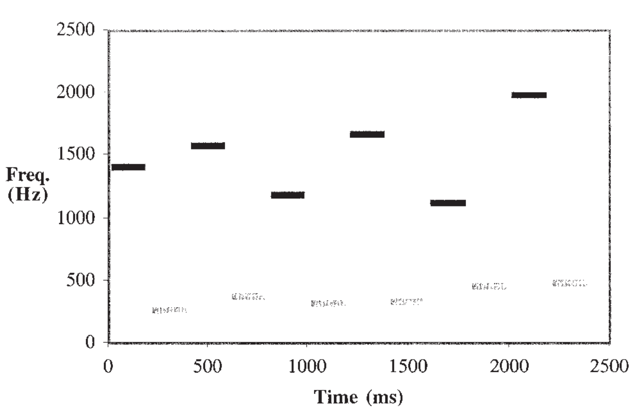

Figure 4. Frequency × Time graph of Experiment 2 stimuli. Each bar depicts the frequency (freq.) and duration of a tone. The odd-numbered (high) tones are indicated by black bars; the even-numbered (low) tones are indicated by gray bars.

Method

Participants

Twelve students at the University of Rochester participated in this experiment. They were screened for the absence of recent music experience as in Experiment 1 and had no reported history of hearing difficulties. All participants received $7.50 for participation.

Stimuli

The sound stimuli were generated in SoundEdit 16.v2, were recorded directly to Sony minidisc, and were presented to individual participants via Koss TD61 stereo headphones as in Experiment 1.

The triplets in this experiment, displayed in bold in the middle section of Table 1, were identical to those in Experiment 1 in every respect except for pitch (see Figure 4). In this experiment, the "odd" set of tones was two octaves higher than the odd set in Experiment 1, whereas the "even" set remained in the same pitch range. Thus, the odd (high) and even (low) tones were separated by roughly two octaves on average.

Exposure. The exposure sequence was identical in structure to the exposure sequence of Experiment 1, with the exception that the two (even and odd) sets of tones were differentiated by a pitch cue.

Test. Test items were identical to those in Experiment 1, with the exception of the heightened pitch of the odd set of tones. Test items are listed in the middle third of Table 1.

Procedure

The methods and instructions for exposure and testing were the same as in the preceding experiment.

Results

Figure 5 presents the mean scores on nonadjacent and adjacent test items in the present experiment, along with the results on the same measures from Experiment 1 for comparison. A 2 × 2 repeated-measures ANOVA was conducted on the data, with Test Item Type (adjacent or nonadjacent) as a within-participants factor and Test Order (adjacent first or nonadjacent first) as a between-participants factor.

There was no effect of Test Order, $F(1, 10) = .53$, $p = .48$, but there was a significant effect of Test Item Type, such that nonadjacent scores ($M = 71.9\%$, $SE = 4.9\%$) exceeded adjacent scores ($M = 47.4\%$, $SE = 3.1\%$), $F(1, 10) = 16.71$, $p = .002$. The interaction term was not significant, $F(1, 10) = 3.34$, $p = .10$. Further, t tests comparing the means for each test item type to chance (50% correct) revealed that participants exceeded chance only on the nonadjacent test, $t(11) = 4.47$, $p = .001$. These results contrast markedly with those of Experiment 1, which showed the opposite pattern. An ANOVA on the data from Experiments 1 and 2 together confirmed that the Experiment (1 vs. 2) × Test Item Type (nonadjacent vs. adjacent) interaction was highly significant, $F(1, 22) = 22.73$, $p < .0001$.

Discussion

In contrast to the results of Experiment 1 with both sets of tones in the same octave, in the present experiment with sets of tones in

[4] We are grateful to Robert Rescorla for drawing this literature to our attention.

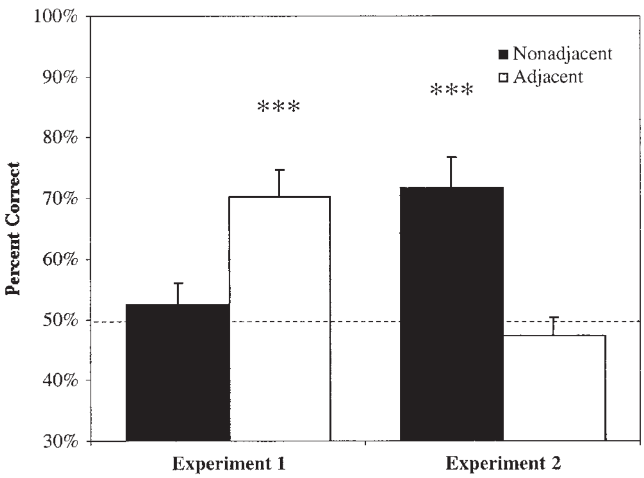

Figure 5. Experiment 2 test scores by item type, with Experiment 1 scores provided for comparison. Error bars are standard errors. $^{***}p \leq .001$.

different octaves, the nonadjacent relationships were readily learned, whereas adjacent relationships were not. Indeed, the results of Experiment 1 are dramatically reversed in Experiment 2 and are in accord with the results of Newport and Aslin (2004) with speech stimuli. Although nonadjacent patterns are typically difficult to learn, they can be acquired when nonadjacent elements fall into two perceptually distinct sets. Taken together, these results suggest that temporal adjacency is not the only constraint that governs statistical learning and that element similarity can interact with temporal adjacency in determining what types of patterned regularities can be acquired. In the two final experiments of this article, we ask whether these results can be replicated and extended to another perceptual cue—timbre.

Experiment 3

The results of Experiments 1 and 2 showed quite clearly that participants have great difficulty acquiring nonadjacent statistics when the pitch register of the temporally adjacent tones is not differentiated but learn nonadjacent statistics readily with such a pitch cue. Experiment 3 was identical to Experiment 2, except that the interleaved tone sets came from the same pitch register (as in Experiment 1) but differed in timbre. If any auditory grouping cue is sufficient to enable the extraction of temporally nonadjacent statistics, then the results of Experiment 3 should mirror those of Experiment 2.

Method

Participants

Twelve students at the University of Rochester participated in this experiment. They were screened for the absence of recent music experience as in Experiments 1 and 2 and had no reported history of hearing difficulties. All participants received $7.50 for participation.

Stimuli

The sound stimuli were generated in SoundEdit 16.v2, were recorded directly to Sony minidisc, and were presented to individual participants via Koss TD61 stereo headphones as in Experiments 1 and 2. To create the complex tones used in this timbre condition, the first five partials[5] of each of the tones used in Experiment 1 were calculated. Triplets were then created so that the odd set tones consisted of Partials 1 and 5 and the even set consisted of Partials 2, 3, and 4, as represented in Table 2. Thus, neither stream was definably "higher" or "lower" in pitch than the other, yet their component frequencies were not identical, leading to the percept of two differing timbres. The three exposure blocks constructed were in other respects identical to the exposure blocks for the first two experiments.

Exposure. The exposure sequence was identical in structure to the exposure sequence of Experiments 1 and 2, with the two timbres in the present experiment substituting for the two frequency ranges in Experiment 2.

Test. Test construction and procedures in this experiment were identical in all respects to those in Experiments 1 and 2, except that sine tones in two different pitch ranges (Experiment 2) were replaced with complex tones of two different timbres. Test items are depicted in the lower third of Table 1.

Procedure

The methods and instructions for exposure and testing were the same as in the preceding experiments.

Results

Figure 6 presents the mean scores on nonadjacent and adjacent test items in the present experiment. A two-way repeated measures ANOVA was conducted, with Test Item Type (adjacent or nonadjacent) as a within-participants factor and Test Order (adjacent first or nonadjacent first) as a between-participants factor. There was a main effect of Test Order, $F(1, 10) = 5.24$, $p = .045$, such that participants receiving the nonadjacent items second scored better overall ($M = 66.14\%$, $SE = 3.0\%$, vs. $M = 55.21\%$, $SE = 3.7\%$). More importantly, there was a main effect of Test Item Type, $F(1, 10) = 9.58$, $p = .01$, and t tests revealed that only the nonadjacent scores exceeded chance ($M = 71.9\%$, $SE = 4.3\%$), $t(11) = 5.07$, $p = .0004$. Scores on the adjacent items did not exceed chance, $M = 49.48\%$, $SE = 4.6\%$, $t(11) = -.11$, $p = .91$.

Discussion

The results of this experiment show that timbre similarity produces parallel effects to that of pitch proximity on statistical learning. When the tones within patterned triplets were presented in the same timbre (and differed in timbre from the intervening tones), their statistical regularities were successfully acquired, even though the tones making up these triplets were not temporally adjacent. In contrast, the moderate-sized (transitional probabilities of .5) statistical regularities among temporally adjacent tones in different timbres were not acquired. The control comparison, in

[5] The term *partial* refers to an integer multiple of a fundamental frequency, with the fundamental frequency itself being the first partial. For instance, for a fundamental frequency of 150 Hz, the first three partials would be 150 Hz, 300 Hz, and 450 Hz. When partials of the same fundamental frequency are presented concurrently, they are perceived as a rich tone at the frequency of the fundamental. This is the case even when the fundamental itself is absent, and the greatest common factor of the partials present is the fundamental (this is called periodicity pitch).

Table 2
Component Frequencies (Partials) of Experiment 3 Stimuli (Hz)

Partial	F_4	G_4	D_4	$G^{\#}_4$	$C^{\#}_4$	B_4
Component frequencies of the odd tones						
1st	349.2	392.0	293.7	415.3	277.2	493.9
5th	1746.0	1960.0	1468.5	2076.5	1386.0	2469.5

Partial	C_4	$F^{\#}_4$	$D^{\#}_4$	E_4	A_4	$A^{\#}_4$
Component frequencies of the even tones						
f_0[a]	[261.6]	[370.0]	[311.1]	[329.6]	[440.0]	[466.2]
2nd	523.2	740.0	622.2	659.2	880.0	932.4
3rd	784.8	1110.0	933.3	988.8	1320.0	1398.6
4th	1046.4	1480.0	1244.4	1318.4	1760.0	1864.8

[a] This is the periodicity of the waveform, but the actual frequency component is not present in the stimuli. This is signified by frequencies in brackets.

which all of these patterns occurred within a single timbre and pitch (Experiment 1), found the opposite results. Taken together, the data show that temporal adjacency and timbre similarity interact in the same way as do temporal adjacency and pitch proximity (Experiment 2). More generally, they suggest that Gestalt-like grouping principles may play an important role in statistical learning, with temporal adjacency interacting with other principles of grouping or similarity to influence which patterns are relatively easy or difficult to learn.

One question that arises from these results, however, is whether these selectivities of learning adjacent versus nonadjacent patterns are merely the result of low-level perceptual constraints, blocking or altering the extraction of temporal order at a very early stage of processing (as would be suggested by a "streaming" account; see Bregman, 1990). Alternatively, and perhaps more interesting, they could be the result of higher-level Gestalt grouping, which might bias learning toward similar elements but not entirely block the perception of actual temporal order. An important difference between these two accounts concerns whether the relevant grouping occurs as a bias on learning itself or rather is a low-level and relatively unalterable sensory or perceptual separation. Our results thus far (only adjacent learning in Experiment 1, only nonadjacent learning in Experiments 2 and 3) are consonant with both explanations. Both the Gestalt literature and the auditory streaming literature have focused on immediate perception, with little research examining whether these perceptual biases affect learning (though see Köhler, 1941; Rescorla & Furrow, 1977; Rescorla & Gillan, 1980; Warren, 1974). Thus, either type of result would contribute to our understanding of sequence learning. However, to understand the mechanisms by which this learning occurs, we sought to determine whether learning is always limited to one or another of these types (adjacent or nonadjacent) or under some circumstances can encompass both. We address this question in our final experiment.

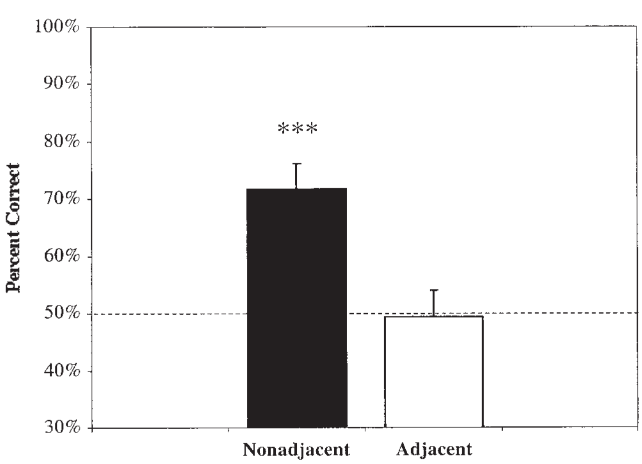

Figure 6. Experiment 3 test scores by item type. Error bars are standard errors. $^{***}p \leq .001$.

Experiment 4

In the preceding three experiments, participants learned only one type of statistical regularity and performed at chance on the other. We have previously suggested that these results may arise from the effects of Gestalt-like grouping or similarity principles on statistical learning: All other things being equal, temporal adjacency tends to dominate, but nonadjacent patterns can readily be acquired if nonadjacent elements are perceptually more similar than adjacent elements (see Newport & Aslin, 2004). However, our results thus far are also amenable to a low-level perceptual account, as postulated by Bregman (1990), who suggested that pitch and timbre cues can lead sounds to group together into perceptual streams, actually blocking temporal adjacency information.

In the present experiment, we examined the mechanism for nonadjacent sequence learning by diminishing the perceptual difference between the two sets of tones, using a timbre contrast that is less extreme than that in Experiment 3. In Experiment 3, the tones were comprised of different partials (Partials 1 and 5 vs. Partials 2, 3, and 4) that overlapped in frequency range. This resulted in their sounding like two different timbres without one stream sounding definably higher or lower than the other. In the present experiment, we used synthetic flute and violin timbres that possessed overlapping spectral profiles, including overlap in the ranges of their fundamental frequencies, and thus sounded more similar to one another than did the timbres used in Experiment 3 while still noticeably differing in timbre. If we could show that, at least under certain circumstances, both the adjacent and nonadjacent patterns can be learned, then we would have evidence that statistical learning can be affected simultaneously by two different biases—adjacency and grouping—rather than the complete dominance of one bias over the other (i.e., streaming). Because of the fact that there were two sets of statistics (adjacent and nonadjacent) that could potentially be learned, one might well expect that the learning of each set of statistics would not be as robust given the same amount of exposure during the learning phase. Therefore, anticipating a smaller effect size, we included a larger number of participants in this study than the number that proved adequate in the preceding experiments.

Method

Participants

Thirty-nine students at the University of Rochester participated in this experiment. They were screened for the absence of recent music experience as in Experiments 1, 2, and 3 and had no reported history of hearing difficulties. All participants received $7.50 for participation.

Stimuli

All sound stimuli were generated in Digital Performer 2.7 using the flute and violin timbres of an AVMSummit MIDI box with a Kurzweil2000 chip. After exporting sounds to minidisc and reimporting them to a Macintosh G3 iMac, the tones were edited to a length of 200 ms each. The flute timbre was used for the even-set stimuli and the violin timbre for the odd-set stimuli. Sound files were inserted into a PsyScope script (J. D. Cohen, MacWhinney, Flatt, & Provost, 1993) for presentation to and testing of individual participants. As in Experiment 3, neither stream was definably higher or lower than the other, but in the present experiment the component frequencies overlapped. The three exposure blocks constructed were in other respects identical to the exposure blocks for the first three experiments.

Exposure. The exposure sequence was identical in structure to the exposure sequence of Experiment 3, with the two timbres in the present experiment substituting for the two timbres in Experiment 3.

Test. Test construction and procedures in this condition were identical in all respects to those in Experiments 1–3, except for the aforementioned timbre substitutions. Test items are listed for reference in the lower third of Table 1.

Procedure

The methods and instructions for exposure and testing were the same as in the preceding experiments, with the exception that the 32-item 2AFC test was administered via the PsyScope interface (J. D. Cohen et al., 1993).

Results

Figure 7 shows the means for both item types in the current experiment. A two-way repeated-measures ANOVA was conducted on the data, with Test Item Type (adjacent or nonadjacent) as a within-participants factor, and Test Order (adjacent first or nonadjacent first) as a between-participants factor. There was no main effect of Test Item Type, $F(1, 37) = 2.58$, $p = .12$, or of Test

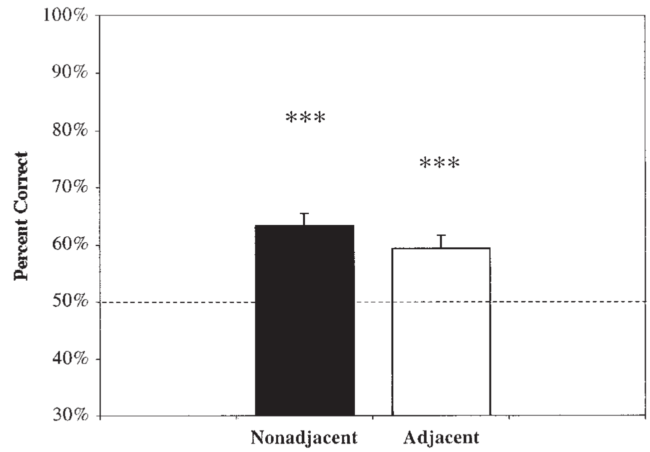

Figure 7. Experiment 4 test scores by item type. Error bars are standard errors. $^{***}p \leq .001$.

Order, $F(1, 37) = 2.10$, $p = .16$, nor was there a Test Item Type × Test Order interaction, $F(1, 37) = 2.24$, $p = .14$.

Collapsed across Test Order, participants scored above chance on both adjacent (M = 63.5%, SE = 2.2%), $t(38) = 6.03$, $p <$ .0001, and nonadjacent (M = 59.5%, SE = 2.3%), $t(38) = 4.14$, $p = .0002$) items. Both of these scores are somewhat lower than the comparable significant effects in earlier experiments.[6] Thus, we examined the possibility that these results represent two subgroups of learners: those who attended to adjacent elements and those who attended to nonadjacent elements. If this is the case, then one would expect a negative correlation between each participant's scores on the adjacent and nonadjacent test items. This possibility was not supported; there was a nonsignificant correlation in the opposite direction, $r(38) = .27$, $p = .09$. Thus, the results of Experiment 4 show that both adjacent and nonadjacent patterns can be learned within the same sequence of tones.

Discussion

The present experiment shows that, if a perceptual cue is moderate in strength, both adjacent and nonadjacent patterns can be learned; therefore, both temporal adjacency and perceptual grouping cues can bias which patterns are learned from a sequence of auditory elements. In our previous work (see Newport & Aslin, 2004), we did not simultaneously, but separately, assess the learning of adjacent versus nonadjacent regularities; thus, we cannot say for certain whether our results using speech stimuli show patterns of learning more like those of Experiments 2 and 3 or those of Experiment 4. However, because natural language users do display knowledge of the relative placement of vowels and consonants in

[6] In an experiment not reported here, we replicated Experiment 2 but replaced one set of tones (even or odd) with a repeating single tone at the average pitch of that tone set. We found that learning was significantly better in this manipulation relative to Experiment 2, suggesting that the presence of multiple simultaneous sequential statistics causes slower learning, perhaps because attention is being focused primarily on one or the other at any given time. We believe that the same effect—slower learning due to multiple statistics—occurred in the current experiment.

syllables (as evidenced, for example, by their ability to learn alphabetic scripts), it would appear that language learners acquire both adjacent and nonadjacent segment regularities, like the learners of Experiment 4 for tone stimuli. The present results also suggest that the ability of a perceptual cue to mark similarity among nonadjacent tones, and thereby lead to the successful learning of nonadjacent regularities, is not due simply to a failure of listeners to hear the temporal order of adjacent elements because of auditory streaming. Rather, our results appear to show a more subtle effect of grouping cues on learning, with the relative strengths of various grouping cues (such as temporal adjacency and timbre) determining the degree to which one dominates over the other in learning.

General Discussion

In this series of four experiments, we have asked several questions about the learning of sequential structure from temporally ordered patterns of tones, and in particular, about how temporal adjacency of related elements interacts with other perceptual grouping cues, such as pitch proximity and timbre similarity. First, are learners able to acquire patterns among nonadjacent elements, particularly if the statistical coherence of these elements is highly regular and consistent? If not, do patterns of statistical information within temporally adjacent elements always take priority over those that are nonadjacent, or does grouping by element properties, such as pitch or timbre, permit learners to acquire statistically regular patterns among temporally nonadjacent elements? More generally, how do temporal adjacency and other perceptual grouping principles affect statistical learning?

Our experiments provide some answers to these questions. In the absence of grouping based on pitch or timbre cues, statistical regularities among temporally adjacent elements take precedence in the statistical computations performed by adult learners, even when they are less reliable than regularities among temporally nonadjacent elements (Experiment 1). However, when a strong perceptual grouping cue was present, whether pitch (Experiment 2) or timbre (Experiment 3), statistics were preferentially computed among elements within the same perceptual group, even when these elements were not temporally adjacent, and were not computed across elements that were temporally adjacent but in different groups. In our final experiment (Experiment 4), we demonstrated that, with a moderate perceptual grouping cue, both adjacent and nonadjacent regularities could be learned successfully (although, in accord with competition between cues, the learning achieved on tests of these two types of regularities was moderate as well). Taken together, these results suggest that statistical learning of tone sequences, and perhaps more generally of auditory element sequences, is constrained by both element adjacency and by perceptual grouping cues.

These results suggest the operation of strong constraints on auditory pattern learning. Temporal adjacency and featural proximity (similarity) have long been known to influence the perception of sequences of elements (Bregman, 1990; Wertheimer, 1923/1938). However, the effects of these factors on pattern learning have not been extensively studied. Our results suggest that these grouping cues can influence the ability of learners to acquire patterns even after fairly extensive exposure. When simple 3-tone melodies were temporally interleaved (and no other grouping cues were provided), 22 min of exposure to four simple melodies (repeated 540 times apiece) resulted in no measurable learning whatsoever. In contrast, when these melodies were differentiated by pitch or timbre, their patterns were readily learned, but their temporal adjacencies were more difficult to acquire. In neither case did learners acquire patterns based simply on the degree of statistical regularity they exhibit. Rather, statistical learning appears to result from the interaction of statistical structure with other factors that affect the tendency to group the elements together.

These findings fit well into a larger literature on constraints on learning. Chomsky (1957, 1965) pointed out that learning is always selective and that certain types of sequential patterns (in his case, human languages) can only be learned successfully by mechanisms exhibiting similar selectivities. In studies of animal learning, Garcia and Koelling (1966), Rescorla and Wagner (1972), and many other subsequent experimenters have shown that even classical conditioning is subject to constraints and sometimes species-specific biases, regarding which stimuli can be readily associated and over how long a temporal interval such associations can occur. Our findings in the present studies suggest that statistical learning, like other types of learning, is not neutral or all-encompassing with regard to the types of regularities it can acquire, but rather is subject to constraints on the types of regularities that can be computed, and on the distance or grouping principles that may interact with such computations.

An important question for continuing research concerns which constraints on learning are widespread and shared across temporally organized domains, like speech and music, and which ones are particular to the domain or type of medium in which the learning occurs. The constraints on auditory grouping exhibited in the present studies—a preference for temporal adjacency but limited or overridden by the similarity of elements in pitch and spectral quality—bear important similarities to other findings on adjacency and nonadjacency, both in speech streams and in other sequentially organized materials.

An extensive literature on visuo-motor sequence learning (e.g., Cleeremans & McClelland, 1991; A. Cohen, Ivry, & Keele, 1990; Hunt & Aslin, 2001; Stadler & Frensch, 1998), as well as our own research on statistical learning of speech and tone streams (Aslin, Saffran, & Newport, 1998; Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran et al., 1999), have shown that human learners are good at learning regularities among temporally adjacent units. However, learning relationships among nonadjacent units is relatively difficult. For example, the results of Gómez (2002), using a simple miniature syntax with only three-word strings, indicate that adult and infant learners acquire relationships between the first and third words of these strings but under only very specific circumstances. Marcus, Vijayan, Bandi Rao, and Vishton (1999), also using three-word strings, obtained learning of both an ABA and ABB regularity in infants, but a replication by Johnson (2002), using temporally ordered visual stimuli, finds the nonadjacent repetition (ABA) to be more difficult to learn. Moreover, in paradigms in which there is a long and continuous stream of elements, rather than only three units in a string, the problem of learning nonadjacent relationships becomes more taxing, and the relevant computations may become even more difficult for learners to conduct (Newport & Aslin, 2004; see also Peña, Bonatti, Nespor, & Mehler, 2002). In the sequence learning literature, Cleeremans and McClelland (1991;

Cleeremans, 1993) have shown that nonadjacent contingencies spanning identical embedded sequences (of three elements or more) are not learned by adults and have suggested that they provide an especially difficult learning problem even for large simple recurrent networks (SRNs).[7]

However, our own recent results on statistical learning of speech streams (Newport & Aslin, 2004) have shown that, although human learners have difficulty with nonadjacent relationships of some types, they are quite capable of learning nonadjacent relationships of other types. In particular, nonadjacent syllable patterns are not readily acquired, whereas patterns among nonadjacent segments of like kinds (consonants or vowels) are learned quite easily. These results are similar to the present findings in which temporal proximity and grouping by category membership also interact. Taken together, these results suggest that learners favor patterns among temporally adjacent units, but that similarity or grouping across elements can override temporal proximity in both speech and tone streams.

In summary, the present results indicate that at least some constraints on statistical learning apply to both speech and other types of temporally ordered patterns. Of course there are important differences between our results with speech streams and those for tone sequences. For example, in our tone sequences, the characteristics that group elements together involve spectral range and quality (that is, pitch and timbre), whereas in our speech sequences, elements group together by virtue of being consonants versus vowels. However, the more general properties of these constraints appear to be shared across domains. At the same time, it is clear that, as patterns become more complex, the constraints on learning that apply to speech and music are likely to differ. In ongoing work, we pursue the question of whether these principles extend to other types of stimuli and where such principles may diverge across domains as patterns become more complex.

[7] Ample evidence exists for learning of higher order information, that is, acquiring information about sequences of Length 3 or greater (e.g., Cleeremans & McClelland, 1991; Curran, 1997; D. V. Howard & Howard, 2001; J. H. Howard & Howard, 1997; Stadler, 1992). However, few studies have attempted to examine nonadjacent dependencies, and those that have done so either had a correlated source of information in an adjacent statistic (J. H. Howard & Howard, 1997) or have failed to find evidence of learning (Cleeremans & McClelland, 1991). For example, J. H. Howard and Howard (1997) presented a A-X-B-X-C-X-D-X pattern and reported evidence of learning after extensive exposure. However, such sequences contain information in adjacent-element strings (AAB, ABB, ACB, and ADB are all more common than AAC, ABC, ACC, and ADC) as well as in nonadjacent strings (A_B), thereby making it unclear whether adjacent or nonadjacent statistics were acquired.

References

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

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology, 44,* 400–413.

Chomsky, N. (1957). *Syntactic structures.* The Hague, the Netherlands: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, the Netherlands: Foris

Chomsky, N. A. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence learning.* Cambridge, MA: MIT Press.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–253.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17–30.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, and Computers, 25,* 257–271.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience, 9,* 522–533.

Dannenbring, G. L., & Bregman, A. S. (1976). Effect of silence between tones on auditory stream segregation. *Journal of the Acoustical Society of America, 59,* 987–989.

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Science, 1,* 327–333.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 458–467.

Garcia, J., & Koelling, R. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4,* 123–124.

Gazdar, G., Klein, E. H., Pullum, G. K., & Sag, I. A. (1985). *Generalized Phrase Structure Grammar.* Oxford, England: Blackwell.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Howard, D. V., & Howard, J. H. (2001). When it *does* hurt to try: Adult age differences in the effects of instructions on implicit pattern learning. *Psychonomic Bulletin and Review, 8,* 798–805.

Howard, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging, 12,* 634–656.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Simultaneous extraction of multiple statistics. *Journal of Experimental Psychology: General, 130,* 658–680.

Johnson, S. P. (2002). Discussion. In N. Z. Kirkham & J. A. Slemmer (Chairs), *New directions in statistical learning.* Symposium conducted at the International Conference on Infant Studies, Toronto, Ontario, Canada.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35–B42.

Köhler, W. (1941). On the nature of associations. *Proceedings of the American Philosophical Society, 84,* 489–502.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1998). Rule learning by seven-month-old infants. *Science, 283,* 77–80.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63,* 176–185.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology, 48,* 127–162.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298,* 604–607.

Pollard, C. J., & Sag, I. A. (1994). *Head-driven phrase structure grammar.* Chicago: University of Chicago Press.

Rescorla, R. A., & Furrow, D. R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 203–215.

Rescorla, R. A., & Gillan, D. J. (1980). An analysis of the facilitative effect of similarity on second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 6,* 339–351.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

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.

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.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development.* New York: Holt, Rinehart & Winston.

Stadler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 318–327.

Stadler, M. A., & Frensch, P. A. (1998). *Handbook of implicit learning.* Thousand Oaks, CA: Sage.

Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. *Cognitive Psychology, 6,* 237–256.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In Ellis, W. (Ed. & Trans.), *A source book of Gestalt psychology* (pp. 71–88). London: Routledge & Kegan Paul. (Original work published 1923).

Wertheimer, M. (1944). Gestalt theory. *Social Research, 11,* 78–99.

Received August 11, 2003
Revision received January 30, 2004
Accepted February 18, 2004 ■