Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 5, 1131–1142

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.5.1131

Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted With Acoustical (Dis)Similarities

Barbara Tillmann
Centre National de Recherche Scientifique, Unité Mixte de Recherche 5020, Université Claude Bernard Lyon 1

Stephen McAdams
Institut de Recherche et Coordination Acoustique-Musique et Centre National de Recherche Scientifique and Ecole Normale Supérieure

The present study investigated the influence of acoustical characteristics on the implicit learning of statistical regularities (transition probabilities) in sequences of musical timbres. The sequences were constructed in such a way that the acoustical dissimilarities between timbres potentially created segmentations that either supported (S1) or contradicted (S2) the statistical regularities or were neutral (S3). In the learning group, participants first listened to the continuous timbre sequence and then had to distinguish statistical units from new units. In comparison to a control group without the exposition phase, no interaction between sequence type and amount of learning was observed: Performance increased by the same amount for the three sequences. In addition, performance reflected an overall preference for acoustically similar timbre units. The present outcome extends previous data from the domain of implicit learning to complex nonverbal auditory material. It further suggests that listeners become sensitive to statistical regularities despite acoustical characteristics in the material that potentially affect grouping.

One fundamental characteristic of the cognitive system is to become sensitive to regularities in the environment via mere exposure to its structure. These implicit learning processes enable the acquisition of highly complex information in an incidental manner and without complete verbalizable knowledge of what has been learned (Reber, 1989; Seger, 1994). Language and music provide two examples of highly structured systems that may be learned in an incidental manner: Native speakers and nonmusician listeners internalize the regularities underlying linguistic and musical structures with apparent ease by mere exposure in everyday life.

Implicit learning processes have been studied in the laboratory with artificial material based on statistical regularities. The material is either created by artificial grammars or based on artificial, simplified language systems. In the seminal studies by Reber (1967), a finite-state grammar was used to generate letter strings with a restricted set of letters. During the first phase of the experiment, participants were asked to memorize the grammatical letter strings but were unaware that any rules existed. During the second phase of the experiment, they were informed that the previously seen sequences were produced by a rule system (which was not described) and were asked to judge the grammaticality of new letter strings. Participants differentiated grammatical letter strings from new ungrammatical ones at better than chance level. Most of them were unable to explain the rules underlying the grammar in free verbal reports (e.g., Altmann, Dienes, & Goode, 1995; Dienes, Broadbent, & Berry, 1991; Reber, 1967, 1989).

In the domain of implicit learning, most research has instantiated the grammars on the basis of visual events (e.g., letters, lights, shapes), and auditory stimuli have rarely been used. Some studies have adapted Reber's artificial grammar design to the auditory domain. The letters of the artificial grammars were replaced by auditory events: sine waves (Altmann et al., 1995), musical timbres (e.g., gong, trumpet, piano, violin, voice in Bigand, Perruchet, & Boyer, 1998), or environmental sounds (e.g., drill, clap, steam in Howard & Ballas, 1980, 1982). In Altmann et al. (1995), for example, letters were translated into tones (i.e., generated with sine waves) by using a random mapping of tone frequencies to letters (e.g., the letter *M* became the musical note *C* with a 256 Hz fundamental frequency), and participants' performance was as high when trained and tested with letter strings as with tone sequences. These studies provided evidence that implicit learning processes also operate on auditory sequences and that the simple exposure to sequences generated by a statistical system allows participants to distinguish sequences that break the rules.

A second set of studies using auditory material used artificial language-like material (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Saffran and collaborators provided evidence for the role of statistical patterns in language acquisition, notably how children learn to

Barbara Tillmann, Centre National de Recherche Scientifique, Unité Mixte de Recherche 5020, Université Claude Bernard Lyon 1, Lyon, France; Stephen McAdams, Sciences et Techniques de la Musique et du Son, Institut de Recherche et Coordination Acoustique-Musique et Centre National de Recherche Scientifique, Paris, France, and Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France.

This research was supported by the Emergence Program of the French Rhône-Alpes Region.

Correspondence concerning this article should be addressed to Barbara Tillmann, Université Claude Bernard—Lyon I, CNRS UMR 5020, Neurosciences et Systèmes, Sensoriels, 50 Avenue Tony Garnier, F-69366 Lyon Cedex 07, France. E-mail: barbara.tillmann@olfac.univ-lyon1.fr

segment the speech flow and to determine beginnings and endings of words. In addition to rhythmic and prosodic cues and to pauses at the end of utterances (Brent & Cartwright, 1996; Jusczyk, Houston, & Newsome, 1999), infants use statistical regularities to discover word boundaries. Saffran et al. (Saffran, Aslin, et al., 1996; Saffran et al., 1999; Saffran, Newport, et al., 1996; Saffran et al., 1997) focused on transition probabilities between syllables that differ inside words and across word boundaries. Transition probabilities take into consideration the co-occurrence between syllables and the absolute frequencies of the syllables. The co-occurrence between syllables leads to greater predictability of word-internal syllable pairs than of syllable pairs spanning word boundaries. In the example *pretty flower*, the syllable *pre* is followed more frequently by *ty* than the syllable *ty* is followed by *flow* because many syllables can follow the word *pretty* but only a few syllables can follow *pre*. In addition, to segment words frequently associated with one another, it is necessary to consider the baseline frequency of syllables in the first position of a pair. For example, when considering that *the* occurs often and is followed by different words, *the sun* is not processed as a unit but segmented into two words. Both types of information thus lead to the statistical cue of transition probabilities,[1] which might be helpful in discovering word boundaries.

On the basis of this rationale, Saffran and colleagues (Saffran, Newport, et al. 1996; Saffran et al., 1997) constructed artificial language-like material as auditory sequences and showed that adults and infants were able to use the statistical regularities to segment the auditory stream. On the basis of 12 syllables, six artificial nonsense words of 3 syllables were created (e.g., bupada, patubi). These words were chained together without pauses or other surface cues in a continuous sequence (e.g., bupadapatubitutibu . . . .). The transition probabilities between 2 syllables inside a word were high (ranging from .31 to 1.00), but the transition probabilities between syllables across word boundaries were weak (ranging from .1 to .2). If listeners were to become sensitive to these statistical regularities, they should be able to extract the words from this artificial language. The experiments consisted of two phases. In a first exposition phase, participants listened to the continuous stream for 21 min (Saffran, Newport, et al. 1996; Saffran et al., 1997) while either being instructed to detect beginnings and endings of words in the nonsense speech (Saffran, Newport, et al., 1996) or to realize an illustration with a coloring program (Saffran et al., 1997). In the second phase of the experiment, participants were tested with a two-alternative forced-choice task: a real word of the artificial language and a nonword (i.e., three syllables that did not create a word of this language and did not occur in the sequence) were presented in pairs, and participants had to indicate the unit that belonged to the previously heard sequence. Participants scored 76% when actively searching for words (Saffran, Newport, et al., 1996) and 59% when doing the coloring task (Saffran et al., 1997, Experiment 1). Repeating the exposition phase increased the performance of participants doing the coloring task to 73% (Saffran et al., 1997, Experiment 2). A more difficult test of participants' learning consisted of contrasting the words with part-words instead of nonwords (Saffran, Newport, et al., 1996). In part-words, two syllables are part of a real word, but the association with the third syllable is illegal within the artificial language. For example, if a legal word is *bupada*, a part-word might contain its first two syllables followed by a different third syllable *bupaka* (with the constraint that this association does not form another word of the artificial language and does not occur over word boundaries in the syllable stream). Even for this test, adult listeners performed above chance. The findings observed for adults have been extended to 8-month-old infants with a simplified language of four words (Saffran, Aslin, et al., 1996). The test phase was based on novelty preferences and the dishabituation effect: Infants' looking times were longer for the loudspeaker emitting nonwords than for the loudspeaker emitting words.

With the goal of showing that the capacity to extract statistical regularities is not restricted to linguistic material, Saffran et al. (1999) replaced the syllables by pure tones at different pitches to create *words of tones* that were chained continuously together to create a sequence.[2] After exposition (i.e., listening three times to the same 7-min sequence), adults and infants performed above chance in the test sessions, choosing between words and either nonwords or part-words. Listeners succeeded in segmenting the tone stream and extracting the tone words as well as they did for linguistic-like sequences of syllables. Overall, the data suggest that statistical learning of syllables and tones can be based on similar knowledge-acquisition processes. This research links studies in the implicit learning domain using artificial grammars to studies concerned with language acquisition (Pacton, Perruchet, Fayol, & Cleeremans, 2001; Winter & Reber, 1994). The extension of conclusions concerning implicit learning derived from artificial laboratory material to more natural material has also been shown for real language (e.g., Mandarin words, Zwisterlood, 1990, reported by Altmann et al., 1995) and new musical systems (e.g., 12-tone music, Bigand, D'Adamo, & Poulin, 2003).

Research on implicit learning in the auditory domain has provided some evidence that listeners become sensitive to regularities underlying both verbal language-like material and nonverbal sounds. For verbal material, the influence of acoustical cues (i.e., stress) on learning has been recently investigated under a developmental perspective (Johnson & Juscyk, 2001; Thiessen & Saffran, 2003, cf. Discussion section). For nonverbal material, up to now, the attribution of the sounds to the statistical regularities has been realized independently of the acoustical characteristics. However, the acoustical features of sounds might have an influence on learning. In Bigand et al. (1998), performance was weaker for musical timbre sequences than for visually presented letters. The authors suggested that the rich acoustical structure of the timbres, and notably the random association of the timbres to the statistical regularities, might have rendered learning of grammatical relations more difficult. A second argument suggesting an influence of acoustical characteristics on statistical learning is based on the example of tonal acculturation. Research in music cognition has provided evidence that nonmusician listeners have acquired so-

[1] The transition probability of B given A is calculated as the frequency of the pair AB divided by the absolute frequency of A (Saffran, Newport, et al., 1996).

[2] The tones were carefully chosen so that the tone words and their chaining in a sequence did not create a context corresponding to the Western tonal music system, and overall, they did not respect tonal rules nor remind listeners of familiar three-tone sequences (i.e., the NBC television network's chimes for American participants).

phisticated knowledge about the Western tonal system by mere exposure to musical pieces obeying its regularities (Francès, 1958/1984; Krumhansl, 1990; Tillmann, Bharucha, & Bigand, 2000). The acoustical structure of complex sounds might make the implicit learning of musical regularities a special case: Some of the statistical regularities are underlined by acoustical characteristics and similarities in the sound. Western tonal music is based on a strong system of statistical regularities, notably regularities concerning the frequencies of co-occurrence between musical events (i.e., notes, chords) and the frequencies of occurrence of musical events (cf. Krumhansl, 1990; Tillmann et al., 2000, for more details). For example, the notes C, E, and G are frequently associated with one another. They all belong to the key of C Major and together they define a C-Major chord. Through mere exposure to musical pieces, listeners become sensitive to these regularities. When hearing the note C, the notes E and G are expected more strongly than are other notes of the chromatic scale (e.g., C# or B). These statistical associations are underlined in some of the cases by the acoustical characteristics of the sound: When the note C is played, the harmonics sounding with the fundamental frequency (at the frequency corresponding to the pitch chroma of C) have frequencies corresponding to the fundamental frequencies of strongly associated notes (i.e., the third and fifth harmonics of the note C correspond to the notes G and E, respectively). In Western tonal music, the rules of the system (creating the statistical regularities) thus coincide at least partially with the acoustical properties of musical sounds. These acoustical features, which support the statistics, might thus help to acquire statistical relations between the tones. This example raises the question about the influence of surface cues (linked to acoustical properties) on learning processes, notably to what extent the acoustical features represent a necessary condition for learning or whether they might facilitate acquisition.

The goal of the present study was to investigate the extent to which acoustical similarities between sounds influence the learning of statistical regularities in nonverbal auditory material. Listeners' performance was compared across situations in which acoustical surface cues either reinforced the statistical structures, contradicted the statistics, or were neutral with respect to them. According to the Gestalt principle of similarity (e.g., Koffka, 1935), similar sounds are grouped together and are segmented into chunks bordered by acoustical dissimilarities. These perceptual units might define attentional units, with the contained information being processed together in memory. The perceptual segmentations based on acoustic similarity, just like segmentations based on temporal proximity or spatial contiguity, might guide attention and influence learning (Mackintosh, 1975; Perruchet & Vinter, 1998, 2002). Statistical relations and associations between sounds might thus be analyzed and learned in a more privileged way inside these units than between sounds spanning unit boundaries. To manipulate acoustical similarities between auditory events and the resulting perceptual segmentations, we used musical timbres selected from the timbre space defined by McAdams, Winsberg, Donnadieu, et al. (1995). Timbre is a multidimensional set of auditory attributes that is based on temporal and spectral features of sounds (cf. Grey, 1977; Krumhansl, 1989; Samson, Zatorre, & Ramsay, 1997). On the basis of dissimilarity judgments, a multidimensional analysis proposed a three-dimensional spatial structure in which 18 synthesized timbres were placed, and distances between timbres reflected perceived dissimilarities among them. For example, the horn timbre is close in space to the trombone timbre (both brass instruments) but is distant from that of the vibraphone (a percussion instrument). Acoustical similarities between timbres influence sequence perception, as reflected in auditory streaming (Bey & McAdams, 2003; Cusack & Roberts, 2000; Gregory, 1994; Iverson, 1995; Singh & Bregman, 1997) and grouping (Deliège, 1987). For example, melody recognition in interleaved melodies increases with increasing timbral dissimilarity, and the segmentation of a musical sequence into groups can also be induced by changes in timbre.

The statistical regularities used in the present study are based on the manipulations of Saffran and collaborators (Saffran et al., 1999; Saffran, Newport, et al., 1996). Groups of three timbres were created to define *statistical timbre triplets.* The experiments are based on the same paradigm that had been used previously with syllables and tones. Participants were first exposed to a continuous stream of timbres based on six timbre triplets and were then tested in a two-alternative forced-choice paradigm to select the statistical triplets against nontriplets (Experiment 1) and part-triplets (Experiment 2).

The crucial point of our study was to systematically manipulate the musical timbres chosen to define the triplets and the boundaries between triplets. Three sequences with six timbre triplets each were constructed on the basis of the distances between timbres in the timbre space. In Sequence 1 (S1), the statistical characteristics were underlined by the acoustical similarities between the timbres: The timbres of the triplets were chosen in such a way that the distances between adjacent timbres inside the triplets were small, but the distances between the last timbre of any given triplet and the first timbre of all other triplets (across boundaries) in the sequence were large. Consequently, the boundaries between triplets were indicated not only by weaker transition probabilities, but also by greater acoustical dissimilarities, that is, when a new triplet starts, the used timbres switched to another part of the timbre space. In Sequence 2 (S2), the distances between timbres inside the triplets were large, but the distances between timbres of two successive triplets (across boundaries) were small. Consequently, the three timbres of a triplet contained two large trajectories in the timbre space but the last timbre of a triplet and the first timbre of the following triplet were acoustically more similar. This similarity might thus "camouflage" the statistical boundary between triplets. In Sequence 3 (S3), the acoustical similarities were neutral with regard to the statistical regularities. Distances were not systematically attributed: Mean distances between timbres inside the triplets were equal to mean distances between timbres of two successive triplets. In S1, the triplets were thus defined by statistical cues and by abrupt acoustical changes between triplets. In S2 and S3, the triplets were only defined by statistical cues, but in S2 the acoustical similarities were out of phase with the statistical boundaries. For the three sequences, the transition probabilities inside the triplets and across triplet boundaries were identical, and the same set of timbres was used. The different attributions of acoustical similarities between timbres sought to investigate the influence of surface cues on the extraction of the statistically recurring units: the triplets. The acoustical surface characteristics might guide the perceptual segmentation and the parsing of the input stream, which might then influence the learning of the statistical triplets.

The systematic attribution of timbres as a function of their distances in the timbre space imposed strong constraints on the constructed sequences. It was not possible to create a second exemplar for each sequence type, which would allow that the statistical triplets of one sequence exemplar could serve as test items for the other sequence exemplar (and vice versa) as in Saffran et al.'s (1999) study on tone sequences. As the statistical triplets thus differed between the three sequences, control groups judged the pairs of triplets in the test phase without having been exposed to the timbre sequence. These control groups allowed us to investigate a general bias in judging triplets that differed in their acoustical structure and to compare the performance of the learning group with this base performance level. In Experiment 1, the three sequences were tested with nontriplets in the test phase.

Experiment 1

Method

Participants

Seventy-two students from the Université de Lyon 1 participated in this experiment.

Stimuli

Definition of the triplets. On the basis of the distances between all 18 synthetic timbres used in McAdams et al. (1995), a subset of 13 timbres was chosen within which the triplets were defined: 1 (French horn), 2 (trumpet), 3 (trombone), 4 (harp), 7 (vibraphone), 8 (striano—a hybrid of bowed string and piano), 9 (harpsichord), 10 (English horn), 11 (bassoon), 12 (clarinet), 13 (vibrone—a hybrid of vibraphone and trombone), 15 (guitar), and 18 (guitarinet—a hybrid of guitar and clarinet).[3] These sounds were all produced with a constant pitch (Eb4, a fundamental frequency of 311 Hz) and a duration of 500 ms. This selected set allowed us to maximize small and large distances between timbres inside triplets and across boundaries for S1 and S2. The 13 timbres were used for the construction of the three sequences S1, S2, and S3. For each sequence, six triplets were defined, with five timbres occurring twice (cf. the Appendix). Table 1 presents the mean distances between timbres inside triplets and between triplets. For S1, timbres were close to each other inside triplets and distant between triplets. For S2, timbres were distant from one another inside triplets, but were close across triplet boundaries. For S3, the mean distances between timbres inside triplets were comparable with mean distances between timbres across boundaries. There was no overlap between the largest distance inside triplets and the smallest distance between triplets for S1 and between the smallest distance inside triplets and the largest distance between triplets for S2. For S3, however, the distances inside and between triplets almost completely overlapped. The differences between means inside triplets and across boundaries were 2.7, 2.9, and 0.3 for S1, S2, and S3, respectively.[4]

Chaining of the triplets. Two hundred twenty exemplars of each triplet were concatenated without silence in random order, with the restriction that the same triplet never occur twice in a row. This sequence stream was divided into three sections of 440 triplets. The transition probabilities between triplets ranged from .1 to .4 ($M = .3$) and within triplets from .5 to 1.0 (mean of .8).

Test triplets. For the test phase, six nontriplets were constructed with the same 13 timbres. As in Saffran et al. (1999) and Saffran, Newport, et al. (1996, 1997), the timbres of nontriplets never occurred in that order in the sequence (even not across boundaries). The mean distances between timbres inside the nontriplets were 5.1, 5.6, and 5.9 for S1, S2, and S3, respectively. The distances ranged from 2.3 to 7.3 for S1, from 2.4 to 8.7 for S2, and from 3.9 to 7.5 for S3. For S1 and S2 (i.e., the two sequences with systematically attributed timbre distances), nontriplets with different internal distances were defined because it was not justified to make the hypothesis that listeners perform only segmentations corresponding to statistical triplets (i.e., groups of three timbres containing two small distances for S1 and two large distances for S2). Three types of nontriplets were created, each instantiated by two nontriplets. Two types of nontriplets contained distance combinations that occurred in the long timbre sequence (referred to as *IN*). One type of nontriplet imitated the distances inside the statistical triplets (referred to as *IN-same*) and contained two small distances for S1 and two large distances for S2. The second type of nontriplet imitated distances of the sequence that also included the covering of a boundary between two triplets (referred to as *IN-mixed*): They contained one small and one large distance in both orders for S1 and S2. The third type of nontriplet contained a distance combination that did not occur in the timbre sequence (referred to as *OUT*): They contained two large distances for S1 and two small distances for S2. If listeners do not learn fine differentiations between timbres, they should be able to reject only these last nontriplets in the test phase. In sum, for both S1 and S2, four nontriplets contained distance combinations that had occurred in the sequence stream, and two nontriplets contained distance combinations that had not occurred before.

Table 1
Description of the Statistical Triplets of the Three Sequences in Terms of Distances in the Timbre Space by McAdams et al. (1995)

	Statistical triplets		
Distances	S1	S2	S3
Intratriplets			
Mean	3.6	6.8	5.9
Min	2.8	5.4	4.5
Max	4.2	8.2	7.3
Intertriplets			
Mean	6.3	3.9	6.3
Min	4.4	3.1	3.9
Max	8.3	4.9	8.7

Note. Distances were based on the CLASCAL algorithm (Winsberg & DeSoete, 1993) applied to dissimilarity ratings on a 9-point scale (cf. McAdams et al., 1995, for details). S = sequence; Min = minimum; Max = maximum.

Apparatus

The timbres were synthesized on a Yamaha TX802 FM Tone Generator as described in McAdams et al. (1995). Timbres were recorded with digital sound software SoundEdit 16 (MacroMedia, San Francisco). The experi-

[3] The numbers refer to the timbres in McAdams et al. (1995, Table 1).

[4] All reported distances were based on distances in the three-dimensional (3-D) space of McAdams et al. (1995). Post-hoc analyses on the chosen triplets provided evidence that these 3-D distances combined the acoustical characteristics of the three dimensions of the timbre space (i.e., attack time, spectral centroid, and spectral flux). For S1, S2, and S3, the distances between timbres inside the triplets and over triplet boundaries were calculated separately for each of the three dimensions. The *t* tests that compared intra- and intertriplet distances for each of the three dimensions taken separately were not significant for any of the sequences (all $ps > .097$ with Bonferroni correction), except for attack time in S2, with greater distances within triplets than between triplets ($p < .001$ with Bonferroni correction).

mental session was run with PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). The stimuli were played at a comfortable listening level via the soundcard of a Macintosh computer.

Design

The between-subjects factors were Sequence Type (S1, S2, S3) and Group (learning, control).

Procedure

In the learning group, participants were exposed to the timbre sequence that was presented in three parts of about 11 min, interleaved with breaks of 1 min. They were asked to simply listen to the sequence without consciously analyzing it but without ignoring it. They were informed that they would be tested afterward on this sequence but were not told about which aspect. At the end of the exposition, they were told that inside the long sequence there existed subgroups of three sounds that formed units (i.e., just like a word forms a unit in language). In the test phase, 36 test pairs (the six statistical triplets were paired with each of the six nontriplets) were presented in a two alternative forced-choice paradigm. The first excerpt was followed by a silence of 500 ms, then the second excerpt was presented. Participants were asked to indicate the excerpt that was part of the previously heard sequence, and if they were uncertain, they were encouraged to choose the excerpt that sounded more familiar considering the timbre sequence heard previously. They answered by pressing one of two keys on the computer keyboard. The test phase started with one practice trial to clarify the structure of the test. The practice excerpts were made out of timbres that did not belong to the sequence, and participants were told that this trial did not have a correct answer.

In the control group, only the test phase was run but with modified instructions. Participants were told that the experimenter's objective was to create a long sequence of sounds in which groups of three sounds should form units (i.e., again explained in comparison to language). Participants were asked to choose the excerpt of the pair that—according to her or him—sounded like a unit and might create a unit when it would be placed in a longer sequence. For the three sequences and for learning and control groups, the test pairs were presented in random order for each participant. Among the 36 pairs, the statistical triplets were presented in the first position for 18 pairs, and the association of the different nontriplets was varied over the pairs. The order of the test items in the pairs was reversed for half of the participants. These presentations were kept constant for the three sequences in learning and control groups.

Results

For learning and control groups, choosing a statistical triplet over a nontriplet was coded as a correct response. Percentages of correct responses (Figure 1, left) were analyzed by a two-way analysis of variance (ANOVA) with Sequence (S1, S2, S3) and Group (learning, control) as between-subjects factors. The main effect of Sequence was significant, $F(2, 66) = 30.29$, $MSE = 121.28$, $p < .0001$. Planned comparisons indicated that correct responses were more numerous for S1 than for S3, $F(1, 66) = 19.73$, and for S3 than for S2, $F(1, 66) = 10.98$ ($p < .01$ with Boole-Bonferroni correction). The main effect of Group was also significant, $F(1, 66) = 30.24$, $MSE = 121.28$, $p < .0001$. Performance was higher for the learning group than for the control group, and this difference was statistically significant for each of the three sequences, $F(1, 66) = 12.99$, $p < .001$, for S1; $F(1, 66) = 8.61$, $p < .01$, for S2; and $F(1, 66) = 8.92$, $p < .01$ for S3. The interaction between the two factors was not significant ($F < 1$), indicating that the differences between control and learning group did not change over the three sequences.

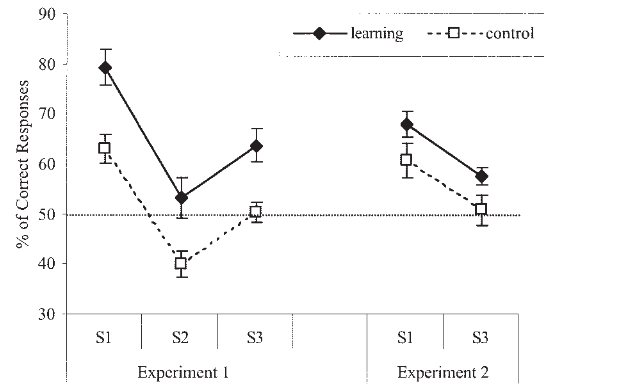

Figure 1. Percentages of correct responses as a function of Group (learning, control) and Sequence (S1, S2, S3) using nontriplets (Experiment 1, left), and part-triplets (Experiment 2, right). For learning and control groups, choosing a statistical triplet over a nontriplet or a part-triplet is coded as a correct response. In S1, statistical regularities were supported by acoustical similarities. In S2, they were contradicted by acoustical properties. In S3, acoustical similarities were neutral with regard to statistics. Chance performance is at 50%.

For S1, performance was different from chance (50%) in both learning and control groups, $t(11) = 8.00$, $p < .001$ and $t(11) = 4.45$, $p < .001$. For S2, performance was not different from chance for the learning group, $t(11) < 1$, but it was less than chance for the control group, $t(11) = -4.0$, $p < .01$. This low percentage (40%) reflects a preference of nontriplets over statistical triplets. For S3, performance was significantly above chance for the learning group, $t(11) = 4.04$, $p < .01$, but not for the control group, $t(11) < 1$.

For S1 and S2, additional analyses separated error rates for nontriplets (i.e., erroneously choosing the nontriplet over the statistical triplet). A first ANOVA separated nontriplets that contained distance patterns appearing in the exposition sequence (IN) from nontriplets that contained distance patterns that did not appear in the sequence (OUT). If test performance of the learning group only reflects the distinction between timbre distance patterns that did or did not occur in the exposition sequence, lower error rates should be observed for nontriplets containing new distance combinations than for the other nontriplets. This pattern would be reflected in a two-way interaction between Type of Nontriplet and Group. However, this pattern was not found ($F < 1$) in a three-way ANOVA with Sequence (S1, S3) and Group (learning, control) as between-subjects factors and Type of Nontriplet (IN, OUT) as within-subjects factor. This analysis confirmed the main effects of Sequence, $F(1,44) = 66.02$, $p < .0001$, and Group, $F(1, 44) = 18.98$, $p < .0001$. Type of Nontriplet interacted only with Sequence, $F(1, 44) = 9.10$, $p < .01$, reflecting a general bias to reject large-distance items in S1 for both learning and control groups. More importantly, the Type of Nontriplet did not interact with Group nor with Group and Sequence (Fs < 1). These results suggest that performance was not a simple function of noting which distance combinations occurred or did not occur in the corpus of triplets versus the entire sequence.

A second ANOVA separated error rates for the three types of nontriplets (see Method) for S1 and S2, with Sequence (S1, S2) and Group (learning, control) as between-subjects factors and Type of Nontriplet (OUT, IN-mixed, IN-same) as within-subjects factor. This analysis confirmed the main effects of both Group and Sequence, $F(1, 44) = 19.71$, $MSE = 403.00$, $p < .0001$; and $F(1, 44) = 54.98$, $p < .0001$, respectively, and the interaction between Type of Nontriplet and Sequence, $F(2, 88) = 3.90$, $p < .05$. This interaction indicated that for S1 (for control and learning groups), the nontriplets with two large distances were rejected more often than the other nontriplets. Again, Type of Nontriplet was not involved in a two-way interaction with Group, nor in a three-way interaction with Group and Sequence (Fs < 1). For nontriplets imitating the distances of statistical triplets (IN-same), the difference between learning and control seemed to be stronger for S1 than for S2, but an additional comparison performed on this specific interaction was not significant, $F(1, 44) = 2.08$, $p = .16$.

Discussion

In Experiment 1, participants were exposed to three types of timbre sequences based on the same statistical regularities but differed in the correspondence between acoustical cues and these regularities. The comparison of the learning groups to the control groups showed that for all three sequences (S1, S2, S3), participants became sensitive to the regularities underlying the timbre sequences to which they had been exposed first. Independent of the acoustical cues, performance increased by 14% on average after the exposition phase. The data of the control group showed that the acoustical similarities influenced the judgments of the listeners. It is interesting to note that the results suggest a preference for triplets with small distances between the timbres. However, this influence of acoustical similarity did not interact with learning for any of the three sequences.

In S3, the timbres were chosen without systematically reinforcing or contradicting the statistical cues. The control group performance showed no preference for the statistical triplets over the nontriplets. The learning group, however, chose the statistical triplets significantly above chance and the control group. The data of the S3 condition thus provide evidence that listeners become sensitive to the statistical regularities inherent in a complex acoustical material, the musical timbre sequences.

For S1, the control group showed a preference for the statistical triplets without exposition to the timbre sequence. The percentage choice of statistical triplets increased significantly after having listened to the timbre sequence with its regularities. For S2, the control group showed an avoidance of statistical triplets, but the percentage choice of statistical triplets also increased after exposition in the learning group and notably by the same amount as for S1 and S3.

The same difference in performance between control and learning groups was observed for all three sequences. Even if acoustical cues indicate different segmentations than do statistical cues in S2, participants picked up some of the regularities that allowed them to overcome their initial response bias and to change their judgments in favor of statistical triplets. However, in the learning group, the performance remained at chance level. Redington and Chater (1996) proposed that two criteria should be fulfilled to safely conclude that learning has occurred: Performance should be significantly different from the control group and from chance. The S2 learning group performance failed to reach their second criterion.[5] The present outcome, notably for S1 and S2, shows the importance of the control group in the experimental design. Without the control group, the amount of learning would have been overestimated in S1 and underestimated in S2. In S3, the control group was situated at chance, the implied comparison line that would have been used without control groups.

The methodological importance of control groups has been discussed previously (Dienes & Altmann, 2003; Meulemans, 1998; Reber & Perruchet, 2003; Redington & Chater, 1996). The control groups allowed us to estimate the initial response bias with which participants judge the experimental material. According to Dienes and Altmann (2003), learning "involves the replacement of one set of biases with another, and so it is potentially useful in investigating learning to know what biases subjects start with" (p. 117). On the basis of this definition of learning, the performance pattern of S2 clearly points to learning after the exposition phase. An a priori comparison against chance does not seem to be appropriate in all cases (e.g., for S1 and S2 in our study). A control group with mean performance at chance seems to be the most preferable situation, also indicating that the test material is bias-free. However, performance at chance might also reflect a lack of motivation on the part of participants because they might believe that nothing but random responding is possible in the task. In consequence, because of motivational deficits, control group performance might underestimate possible performance levels. This criticism against the use of control groups was advanced by Redington and Chater (1996) and applies to both trained and untrained control groups. In our study, only the control group tested with S3 was at chance level, and the groups tested with S1 and S2 were above and below chance, respectively. This outcome suggests that participants tried to follow the experimental instructions and were sensitive to the acoustical characteristics of the material.

Previous studies in the implicit learning domain have reported similar cases for control groups, with performance being either significantly above (Dulany, Carlson, & Dewey, 1984; Meulemans & Van der Linden, 1997; Redington & Chater, 1994, quoted by Redington & Chater, 1996) or below chance (Meulemans & Van der Linden, 1997, Experiment 2b). Performance above and below chance suggests that control participants might either learn something about the material during the test phase or base their judgments on characteristics other than statistical ones inherent to the material, for example, "knowledge about what a typical item looks like" (Redington & Chater, 1996, p. 127). In consequence, when some characteristics of the material or participants' interpretation of typicality are positively or negatively correlated with statistical features of the sequences, the performance of control groups might fall above or below chance. Meulemans and Van der Linden (1997, Experiments 2a and 2b) made the link between the associative chunk strengths of letter sequences and participants' preferences for elements with higher associative chunk strength. In their Experiment 2b, for example, the ungrammatical items had a

[5] Three participants had very low performance at 33%, and the remaining participants had a performance level of 60%, which is above chance, two-sided $t(8) = 3.65$, $p < .01$.

higher associative strength, and the control group participants picked up characteristics of the material in order to base their answers on this aspect. In our study, the characteristics of the material were linked to the acoustical features of the timbres inside the triplets. In S3, acoustical similarities were attributed unsystematically, and the chance performance suggests that no particular response bias was present. In S1 and S2, the acoustical similarities were systematically attributed to the triplets, and this manipulation seems to introduce a preference to choose triplets with smaller distances (i.e., S1). The outcome of the control groups suggests that listeners are biased in their judgments, notably in the sense that timbrally similar events are more often judged as forming units than are dissimilar events. Control performance thus reflects participants' response biases based on the perceptual properties of the sounds (i.e., timbral similarity). This bias probably existed before the experiment and seems to be rather general (i.e., it is not limited to timbres but also applies to other events with perceptual similarities). To some extent, this bias might remain in the learning group, but the increase of performance (i.e., choosing statistical triplets more often) shows that the exposition phase had an effect on participants' answers. Independent of perceptual properties and preference biases linked to these perceptual properties, learning of statistical regularities took place in all three sequences.

Experiment 2

Together with Saffran's research, the data of Experiment 1 suggest that adult learners segment sequences of complex auditory information regardless of whether the input is linguistic (syllables), simple nonlinguistic (tones played with sine waves), or complex nonlinguistic (timbres). We further explored in Experiment 2 the statistical learning in timbre sequences by using a more difficult discrimination test following learning. Participants were required to distinguish statistical triplets from triplets containing parts of them (i.e., part-triplets consisted of two timbres occurring in that order in a statistical triplet associated with a third timbre as in Saffran et al., 1999; Saffran, Newport, et al., 1996). This measure provides a stronger test of learning because correct performance requires discriminating two triplets that differ by only one timbre. Experiment 2 thus focused on the comparison between S1 and S3 and investigated whether acoustical similarity reinforcing statistical relations (S1) might help to improve performance in comparison with an acoustically neutral situation (S3). In the studies by Saffran and colleagues (Saffran et al., 1999; Saffran, Newport, et al., 1996), participants succeeded in performing the part-triplet test for both syllable and tone sequences. If the same mechanisms are at work for the timbre sequences, listeners should be able to perform the triplet–part-triplet discrimination in Experiment 2.

Method

Participants

Forty-eight students from the Université de Lyon 1 participated in this experiment. None had participated in Experiment 1.

Materials

Sequences S1 and S3 of Experiment 1 were used. For the test session, six part-triplets were constructed with the timbres of the sequences. As in Saffran et al., 1999, and Saffran, Newport, et al. (1996), two timbres of the part-triplets had occurred in that order in the statistical triplets, but the association with the third timbre had not occurred in the sequence stream. Three part-triplets kept the timbres in Positions 1 and 2 of three statistical triplets, and three part-triplets kept timbres in Positions 2 and 3 of the three remaining statistical triplets. The mean distances for the part-triplets were 4.2 (ranging from 2.3 to 8.1) for S1 and 5.6 (ranging from 4.2 to 7.8) for S3. Two types of part-triplets were constructed for S1 that both contained distances that occurred in the timbre sequence: Three part-triplets contained only small distances (like the two distances inside the statistical triplets, referred to as IN-same), and three contained one small and one large distance (thus including a distance crossing the boundary between triplets, referred to as IN-mixed).

Design and Procedure

The between-subjects factors were Sequence type (S1, S3) and Group (learning, control). The procedure was as described in Experiment 1 for learning and control groups. In the test phase, nontriplets were replaced by part-triplets.

Results

For learning and control conditions, choosing the statistical triplet was defined as a correct response. Percentages of correct responses (Figure 1, right) were analyzed by a two-way ANOVA with Sequence (S1, S3) and Group (learning, control) as between-subjects factors. As in Experiment 1, the main effects of Sequence and Group were significant, $F(1, 44) = 13.84$, $MSE = 89.94$, $p < .01$, and $F(1, 44) = 6.43$, $MSE = 89.94$, $p < .05$, respectively, but the two factors did not interact ($F < 1$). Correct responses were more numerous for S1 than for S3 and for the learning group than for the control group. The difference between the learning and control groups failed to reach significance both for S1, $F(1, 44) = 3.43$, $p = .07$, and for S3, $F(1, 44) = 3.01$, $p = .09$. For S1, performance was different from chance for the learning group, $t(11) = 7.03$, $p < .001$, and the control group, $t(11) = 3.26$, $p < .01$. For S3, performance was different from chance for the learning group, $t(11) = 4.21$, $p < .01$, but not for the control group ($t < 1$).

For S1, percentages of errors (i.e., incorrectly choosing part-triplets over statistical triplets) were separated for the two types of part-triplets and were analyzed by a two-way ANOVA with Group (learning, control) as between-subjects factor and Type of Part-Triplet (IN-same, IN-mixed) as within-subjects factor. The main effect of Group failed to reach significance, $F(1, 22) = 2.90$, $p = .10$. For the control group, percentages of errors were almost identical for the two types of part-triplets, but for the learning group they were lower for IN-same than for IN-mixed part-triplets. However, this interaction was not significant, $F(1, 22) = 1.74$, $p = .20$.

Discussion

Experiment 2 tested the statistical triplets against part-triplets for S1 and S3. The data showed that even with this more difficult test condition, participants succeeded in the task. For S3, the learning group performance was above chance and above the control group. For S1, the learning group performance was above the control group performance that, as in Experiment 1, was also above chance. The more difficult test condition was reflected in a smaller difference between control and learning groups than in Experiment 1. In the test condition using nontriplets, the percentage choice of statistical triplets increased from control group to learning group by 15% (averaged over S1 and S3). In the test condition using part-triplets, this increase was only 7%. In studies by Saffran, Newport, et al. (1996) and Saffran et al. (1999), a comparable decrease in performance was observed between the two test conditions: For syllables, participants performed at 76% for nonwords but at 65% for part-words. For tones, performance was at 77% for nonwords and at 65% for part-words.

Performance of control and learning groups showed that the acoustical similarities induced a preference bias for congruent triplets, leading to generally increased percentages for S1. Concerning the influence of acoustical similarities on learning, the data of Experiment 2 confirmed the outcome of Experiment 1. As in Experiment 1, the amount of learning was reflected in the change between control and learning group. When the acoustical and statistical information were congruent (S1), the amount of learning was not increased in comparison with the situation containing only statistical information (S3). In other words, listeners did not take advantage of a perceptual segmentation that additionally indicated the statistical features with acoustical surface cues.

General Discussion

In the present study, we used musical timbre sequences to investigate the influence of acoustical surface characteristics on the implicit learning of statistical regularities. In line with Saffran's (Saffran et al., 1999; Saffran, Newport, et al., 1996) research, statistical triplets of timbres were chained together in a continuous sequence, and participants were then tested to recognize the statistical triplets against nontriplets or part-triplets. The statistical regularities were based on transition probabilities between timbres in the sequence: These transition probabilities were higher between adjacent timbres inside the statistical triplets than between timbres across triplet boundaries. The contribution of our study was (a) the use of acoustically complex material (i.e., musical timbres) and (b) the systematic manipulation of acoustical similarities between timbres while keeping constant the statistical regularities. The acoustical similarities were based on the distances between timbres in a three-dimensional timbre space (McAdams et al., 1995): The closer the timbres are, the more similar they are perceived to be. It was felt that manipulating timbral similarity would affect perceptual segmentation of the sequences, which in turn would reinforce the statistical structure or not. Three types of timbre sequences were defined: In S1, the statistical regularities were supported by acoustical similarities (i.e., a triplet boundary was indicated by a weak transition probability and a dissimilarity between timbres), in S2, the statistical regularities were contradicted by acoustical similarities (i.e., abrupt acoustical change occurred inside the triplets and more similarly sounding timbres crossed the boundaries), and in S3, the acoustical characteristics were attributed in a neutral way with regard to the statistical regularities. The outcome of Experiment 1 showed that the ability to choose statistical triplets over nontriplets improved in the same way for the three sequence types. Independent of the acoustical similarities, participants became sensitive to the statistical structure of the timbre sequence heard in the exposition phase. This finding was confirmed for S1 and S3 in Experiment 2 with a more difficult test phase opposing triplets and part-triplets. The main finding of our study was that listeners learned the statistical regularities of complex auditory material and that its surface characteristics (which affected grouping and overall preference bias) did not affect this statistical learning.

Learning Statistical Regularities With a Complex, Multidimensional Auditory Material

The overall outcome confirmed the findings of Saffran and collaborators (Saffran et al., 1999; Saffran, Newport, et al., 1996) and extended them to a more complex nonlinguistic material. With simple exposition to a structured auditory material, listeners became sensitive to the statistical regularities inherent in this material, which allowed them to extract subunits of three elements. This learning process thus does not apply only to syllables (Saffran, Newport, et al., 1996; Saffran et al., 1997) and pure tones (Saffran et al., 1999), but also to more complex, nonlinguistic material: musical timbres. Our study therefore provides corroborative evidence to other empirical findings demonstrating that regular structures of an auditory, nonlinguistic environment can be internalized through passive exposure. The previously tested regularities were derived from finite-state grammars initiated with musical timbres (Bigand et al., 1998) and environmental sounds (Howard & Ballas, 1980, 1982) and from the structure of contemporary 12-tone music, which is based on frequency distributions of tone intervals (Bigand et al., 2003). The investigation of statistical learning processes has been extended to artificial syntactical structures based on linguistic sequences (i.e., nonwords like biff, jux, tiz) and nonlinguistic sound sequences (i.e., alert sounds like an ascending buzz or chimes; Saffran, 2002).

A short exposure time to timbre sequences containing transition probability patterns is sufficient to enable participants to segment the continuous stream into units and subsequently to choose statistical triplets over nontriplets or part-triplets. Becoming sensitive to the statistical regularities in the timbre sequence might result from simple exposure without necessarily relying on explicit processes of analysis. The timbres used were not natural instrument sounds but were synthesized, and some of them were hybrids (i.e., they were constructed by combining two different timbres). This kind of material makes it more difficult to explicitly verbalize and learn the timbre combinations than it would have been for letter or syllable strings. General informal comments of participants after the experiment testified to their overall uncertainty in giving their answers. To further investigate this issue, future studies along this line will include confidence ratings after each judgment (Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Berry, 1997). Low or absent correlation between confidence ratings and grammaticality judgments are taken as an indication for the implicit nature of acquired knowledge.

Reber (1992) stated that "implicit learning is the default mode for the acquisition of complex information about the environment" (p. 25). The artificial materials used in the lab are simpler than environmental sequences of events. However, the same basic principles of learning may serve as a model for understanding the implicit learning processes in natural environments (Winter & Reber, 1994). The sounds used here are acoustically more complex than, for example, pure tones, and if we consider timbre to be one of the main perceptual vehicles for sound source identity (McAdams, 1993), studies demonstrating implicit learning of sequential relations among timbres may shed light on the acquisition of knowledge about sequential expectancies and dependencies among sources in the natural environment. The present study is thus part of more recent research representing a further step to bridge the gap between artificial lab material and real world material (Pacton et al., 2001; Saffran, 2002; Winter & Reber, 1994 for language; Bigand et al., 2003; Tillmann et al., 2000 for music).

Acoustical Similarities Create a Preference Bias

The data obtained with the timbre sequences, and notably the performance of the control groups, suggest that acoustical similarities introduce a response bias. Small distances between timbres inside a triplet are preferred over larger distances when judging whether the triplet forms a perceptual unit. This general bias led to performance in the control group that was above chance for S1 and below chance for S2. These differences in response preferences for triplets reflect the influence of the manipulated distances in the timbre space. These data provide further support for the perceptual reality of the musical timbre space as a model of timbre perception that goes beyond the initial similarity judgments used to derive it (McAdams et al., 1995). Previously, timbre space representations have been shown to model several perceptual aspects of timbre. They allow the positioning of new timbres (such as hybrid timbres) in a coherent way (Grey & Gordon, 1978; Krumhansl, 1989; Wessel, 1979), and they predict confusion errors in identification tasks (Grey, 1977), the degree to which alternation between timbres will result in auditory stream segregation (Bey & McAdams, 2003; Iverson, 1995) and the perception of abstract, transposable relations among timbres (Ehresman & Wessel, 1978; McAdams & Cunibile, 1992; Wessel, 1979).

Learning Independently of Acoustical Similarities

Independently of match or mismatch between the perceptual units created by timbral similarities and the statistical units, listeners became sensitive to the statistical regularities. In comparison with the control groups, choosing statistical triplets increased by a similar amount for learning groups of the three sequences. Concerning the influence of acoustical features, S3 can be considered as a baseline condition, because acoustical variations were not systematically related to statistical regularities. For S3, the learning groups chose more often the statistical triplets over test triplets than did the control groups, which were at chance level. The same improvement of performance was observed for the learning groups of S1 and S2. The acoustical features manipulated in and between the triplets represented neither an advantage nor a disadvantage for learning. In other words, the learning of transition probabilities takes place straddling segmentations and groupings induced by acoustic characteristics. Despite the surface-based units, the statistical triplets became salient for the listeners, suggesting that the Gestalt principle of similarity did not restrict the statistical learning in the timbre sequences (even if it led to an overall preference bias).

This outcome points to the strength of the cognitive system in acquiring sensitivity to the underlying regularities of the environment. The present data further suggest that the lower performance with timbres in comparison with letters (Bigand et al., 1998) might not be solely because of the acoustical complexity of the timbre material. The lower performance might also be caused by differences in either modality (visual vs. auditory), presentation (simultaneous vs. sequential), or the fact that letter strings might be more easily coded than timbres in an explicit, verbalized way (at least parts of it in the form of chunks). Concerning the learning of Western tonal music by nonmusician listeners, our data suggest that tonal acculturation might also take place without the convergence between statistical patterns of tone use and acoustical properties. This finding can be interpreted as being rather encouraging for contemporary composers creating new musical systems: Listeners might become sensitive to new systems of regularities instantiated by complex, synthesized sound types via mere exposure. Similar observations can be reported for nonwestern musical systems as, for example, the Arabic music system, in which scale structures do not mirror the structure of complex sounds. The basic Arabic scale (e.g., notably C D E-half-flat F G A B-half-flat C') contains many quarter-tone intervals for which correspondences do not exist between regularities in terms of frequency ratios and statistical use. However, native listeners acquire great sensitivity to the underlying regularities of the musical system (Ayari & McAdams, 2003).

In contrast to our data on musical timbres, some influence of acoustical surface markers on statistical learning has been reported for linguistic material in adults and infants. In the following, we discuss differences between material and methods, and we outline future research exploring potential parameters influencing the respective strengths of acoustical and statistical cues in learning.

In artificial language material, Saffran, Newport, et al. (1996, Experiment 2) manipulated vowel length as one indicator of prosody: Statistical cues were either accompanied by vowel lengthening of one of the syllables (i.e., either the final or initial syllable of the word) or not (i.e., no lengthening was applied).[6] Adult participants became sensitive to the statistical regularities independently of the acoustical markers, although lengthening of the final syllable improved word learning (80%) in comparison with the absence of lengthening (65%) and with lengthening of the initial syllable (61%). The authors suggested that listeners' implicit knowledge about the rhythmic structure of the English language influences the discovery of word units. In several languages, including English, lengthening of the final syllable indicates word endings. Participants might use this knowledge as a strategy to parse the incoming auditory stream. As our timbre material is nonlinguistic, this particular "parsing-strategy" in reference to language might not be as strongly activated as for linguistic-like material. This difference might explain why we observed comparable amounts of learning independent of acoustical surface cues: Performance did not increase more for S1, in which acoustical cues were paired with statistical cues, than for S3 without systematic acoustic pairing. In addition to the difference in material type (linguistic vs. nonlinguistic), acoustical cues in S1 marked beginnings and endings of the triplets. In light of the improved learning due to final vowel lengthening, one might wonder whether—even for the nonlinguistic material—marking solely the ending of the timbre triplets

[6] To avoid the criticism that participants have an initial preference bias for words with a specific lengthening, some of the nonwords also contained lengthening on an initial or final syllable (i.e., no control group was used).

might be more efficient and improve performance in comparison with triplets without surface cues or with markers for both beginnings and endings.

In Saffran, Newport, et al. (1996, Experiment 2), none of the prosodic lengthening cues worked against the statistical regularities. Johnson and Jusczyk (2001) manipulated speech surface cues (i.e., coarticulation or stress) to pit them against statistical cues. In contrast to our study, part-words of the test phase occurred in the exposition sequence: Two syllables were part of a word and the third syllable was associated across a word boundary. As the predominant stress pattern in English is strong–weak (i.e., with the stressed syllable marking the onset of a word), the first syllable of the part-word was marked with a stress (Johnson & Jusczyk, 2001, Experiment 2). For 8-month old infants, these competing acoustical surface cues outweighed statistical cues and reversed infants' listening preferences from statistical words to part-words. When discussing differences between Saffran, Newport, et al.'s (1996) data and their data, the authors emphasized that stress is a stronger indicator than vowel lengthening and further suggested that "infants and adults may weight conflicting segmentation cues differently" (p. 563). When linking this outcome to our data, the question arises concerning the degree to which the observed differences are due to the nature of the material (linguistic-like vs. nonlinguistic) or to developmental changes of the cognitive system. Answers might lie in testing either adults with their language-like material or infants with our timbre material.

Recent data on the influence of stress on statistical learning of linguistic-like material seem to favor a developmental hypothesis (Thiessen & Saffran, 2003). With 9-month old infants, the previously reported advantage of stress cues over statistical cues was replicated for material similar to the one used by Johnson and Jusczyk (2001). However, 7-month old infants did not show this bias for perceptual indicators and became sensitive to the statistical relations between the syllables despite stress cues indicating conflicting groupings. According to the authors, this outcome points to infants' original sensitivity to statistical cues, which allows them to extract first words, and when becoming more sensitive to stress patterns, they are biased to stress patterns of their language. Listener's knowledge about stress patterns of their native language thus influences the learning of artificial language-like material (Johnson & Juscyk, 2001; Saffran, Newport, et al., 1996; Thiessen & Saffran, 2003). It also influences performance in perceptual segmentation tasks applied on a second language, acquired later in life (Sanders, Neville, & Woldorff, 2002). However, if the to-be-learned material does not refer to previous knowledge of segmentational acoustical cues, as in our present timbre study, the learning processes are sensitive to the statistical characteristics of the material, independently of the acoustical indications.

A further line of research investigating the observed differences on the influence of acoustical versus statistical cues consists of applying Johnson and Jusczyk's (2001) manipulation to nonlinguistic material. Stress patterns might be imitated by changes in the salience of musical timbres. Large distances and short attack times are possible candidates allowing the creation of rhythmicity or stressed accent patterns. Future experiments will first determine timbre combinations that create accents for listeners, and these timbres will then be used to indicate stressed events in the triplets (notably on the first timbre). The statistical cues might still overwhelm the acoustical cues, the reverse pattern might be observed, or the conflicting cues could cancel each other out. If the timbre sequences are processed independently from prior knowledge on stress structure in language, statistical cues should still outweigh acoustical cues. If, however, the underlying processes and knowledge-based biases are more domain-general, then the stress cues should overcome or at least disturb the learning of statistical cues for this nonverbal material as well.[7]

Conclusion

Our present study extends Saffran's (Saffran et al., 1999) research on tone sequences to a more complex auditory material (timbres) and shows that learning of statistical regularities is possible independent of their being reinforced or not by acoustical segmentation cues. The same change in performance between control and learning groups was observed even when acoustical features were pitted against statistical relations over triplet boundaries. Even between dissimilar sounds and despite conflicting perceptual groupings, the cognitive system seems to become sensitive to statistical associative relationships. In everyday life, this capacity of the cognitive system seems to be rather useful as, for example, associations and statistical regularities (also concerning the temporal ordering of sounds) have to be learned between complex environmental sounds that can differ acoustically. For environmental sounds, semantic relations and coherence of actions might also play a role in the acquisition process. The results of Howard and Ballas (1980) point to a similar conclusion: Listeners became sensitive to regularities generated by an artificial grammar for both tone sequences and environmental sound sequences, but performance decreased for semantically incoherent environmental sound sequences. In this latter case, the authors suggested that listeners' knowledge about coherent semantic chaining interfered with the new learning. For our timbre sequences, no previously acquired semantic knowledge or other type of knowledge would be implicated either to help or to render more difficult the learning of the new regularities because the listeners would be unlikely to have such knowledge for these sounds and sequences.

[7] For direct comparisons with Johnson and Jusczyk (2001), this new timbre material that included accent patterns should be defined in such a way that both triplets and part-triplets occur in the exposition sequence.

References

Altmann, G. T., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 899–912.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsim). *Music Perception, 21,* 159–216.

Bey, C., & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 267–279.

Bigand, E., D'Adamo, D. A., & Poulin, B. (2003). *The implicit learning of twelve-tone music.* Manuscript submitted for publication.

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive, 17,* 577–600.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 96–125.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments and Computers, 25,* 257–271.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics, 62,* 1112–1120.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception, 4,* 325–359.

Dienes, Z., & Altmann, G. T. M. (2003). Measuring learning using an untrained control group: Comment on R. Reber and P. Perruchet. *The Quarterly Journal of Experimental Psychology, 56A,* 117–123.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Implicit and explicit knowledge bases in artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1322–1338.

Dienes, Z., & Berry, D. C. (1997). Implicit learning: Below the subjective threshold. *Psychological Bulletin and Review, 4,* 3–23.

Dienes, Z., Broadbent, D., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 875–887.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General, 113,* 541–555.

Ehresman, D., & Wessel, D. L. (1978). Perception of timbral analogies. *Ircam Report, 13.*

Francès, R. (1984). *La perception de la musique* [The perception of music] (2nd. ed.). Paris: Vrin. (Original work published 1958)

Gregory, A. H. (1994). Timbre and auditory streaming. *Music Perception, 12,* 161–174.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61,* 1270–1277.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America, 63,* 1493–1500.

Howard, J. H. J., & Ballas, J. A. (1980). Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception & Psychophysics, 28,* 431–439.

Howard, J. H. J., & Ballas, J. A. (1982). Acquisition of acoustic pattern categories by exemplar observation. *Organization, Behavior and Human Performance, 30,* 157–173.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 751–763.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159–207.

Koffka, K. (1935) *Principles of Gestalt psychology.* New York: Harcourt, Brace & World.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–54). Amsterdam: Excerpta Medica.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Mackintosh, N. J. (1975). A theory of attention: Variations of the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford, England: Oxford University Press.

McAdams, S., & Cunibile, J.-C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society Series B, 336,* 383–389.

McAdams, S., Winsberg, S., Donnadieu, S., DeSoete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research, 58,* 177–192.

Meulemans, T. (1998). *L'apprentissge implicite [Implicit Learning].* Marseille, France: Solal.

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1007–1028.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401–426.

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

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences,* 25, 297–330.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.

Reber, A. S. (1992). *Implicit learning and tacit knowledge.* New York: Oxford University Press/Claredon Press.

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *The Quarterly Journal of Experimental Psychology, 56A,* 97–116.

Redington, M., & Chater, N. (1994). The guessing game: A paradigm for artificial grammar learning. In A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th Annual Meeting of the Cognitive Science Society* (pp. 745–749). Hillsdale, NJ: Erlbaum.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General, 125,* 123–138.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47,* 172–196.

Saffran, J., Aslin, R., & Newport, E. (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. *Psychological Science, 8,* 101–105.

Samson, S., Zatorre, R., & Ramsay, J. O. (1997). Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology, 51,* 307–315.

Sanders, L. D., Neville, H. J. & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic and speech-pattern cues. *Journal of Speech, Language and Hearing Research, 45,* 519–530.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115,* 163–169.

Singh, P. G., & Bregman, A. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America, 102,* 1943–1952.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39,* 706–716.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychology Review, 107,* 885–913.
Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3,* 45–52.
Winsberg, S., & DeSoete, G. (1993). A latent class approach to fitting the weighted Euclidean model, clascal. *Psychometrika, 58,* 315–330.
Winter, B., & Reber, A. S. (1994). Implicit learning and the acquisition of natural languages. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 114–145). San Diego, CA: Academic Press.
Zwisterlood, P. (1990, November). *Speech segmentation during the first minutes of second-language acquisition.* Paper presented at the 31st Annual Meeting of the Psychonomic Society, New Orleans, LA.

Appendix

Statistical Triplets of the Three Sequences

S1: 13–4–7; 9–15–7; 11–3–12; 11–2–8; 10–1–12; 10–3–18
S2: 8–9–11; 1–4–10; 2–7–11; 2–13–18; 3–15–10; 3–7–12
S3: 13–18–7; 10–4–3; 9–8–7; 9–2–12; 11–15–3; 11–8–1

Note. Numbers refer to the timbres in McAdams et al. (1995, Table 1): 1-French horn, 2-trumpet, 3-trombone, 4-harp, 7-vibraphone, 8-striano (a hybrid of bowed string and piano), 9-harpsichord, 10-English horn, 11-bassoon, 12-clarinet, 13-vibrone (a hybrid of vibraphone and trombone), 15-guitar, and 18-guitarinet (a hybrid of guitar and clarinet). Note that the names refer to the instrument that the synthetic sound was meant to simulate. S = sequence.

Received August 19, 2003
Revision received February 17, 2004
Accepted February 20, 2004 ■