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 becomeaffect 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 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 im-

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 (78and)-239.3(th5es)-5(betw cs)-5
to three of triplets. The meciration pbabilirances
between31335.9(triplets31335.(megused)-335.3(from)-335..1en31335.9oom)-335..4omst-
ere381.7(sequenis381.(o(onis381.(anis381.1(across381.9(boundariix).381.T(ere381.5(mea

ForandlaromdistancesForthreeoftriplets

thetimbedsequenis

—

triplets4 -15.mbferuused4 -15.9oed4 -15.(wae)]TJ 2F5 1 T-33723562 0 TDIN-4ndar—ixe

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, $F(2, 88) = 3.90$, $p <$

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 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

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

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 ■