Implicit Statistical Learning of Language and Music

Bénédicte Poulin-Charronnat * & Barbara Tillmann *
LEAD – CNRS UMR5022, Université de Bourgogne

ABSTRACT
One fundamental ability of the human cognitive system is to become sensitive to the regularities present in the environment. By mere exposure to the surrounding environment, and without intention to learn, perceivers are able to adapt to the regularities of their environment, whether perceptual, linguistic, musical, motor, or even social. Language and music are two highly structured systems; they convey statistical regularities that can be learned by mere exposure through implicit learning mechanisms. The present article focuses on laboratory studies mainly using behavioral methods to investigate the statistical learning in language and music domains. We sequentially review research showing the structural segmentation of sequential material and the acquisition of grammatical or syntactic structures in artificial systems, followed by extensions to more real-world-like materials. The final section is dedicated to a brief overview of the potential mechanisms underlying this learning, as proposed from different theoretical perspectives.

Keywords: implicit statistical learning, language, music

Apprentissage Statistique Implicite du langage et de la musique

RÉSUMÉ
Une capacité fondamentale du système cognitif humain est de devenir sensible aux régularités présentes dans l'environnement. Par simple exposition au milieu environnant, et sans intention d'apprendre, les individus sont capables de s'adapter aux régularités qui les entourent, qu'elles soient perceptives, linguistiques, musicales, motrices, ou même sociales. Le langage et la musique sont deux systèmes hautement structurés ; ils présentent des régularités statistiques qui peuvent être apprises par simple exposition à travers des mécanismes

* Author note: https://orcid.org/0000-0003-3244-6747, https://orcid.org/0000-0001-9676-5822.
Correspondence: LEAD – CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon. Email: benedicte.poulin@u-bourgogne.fr or barbara.tillmann@cnrs.fr.

d'apprentissage implicite. Le présent article se concentre sur des études conduites en laboratoire, utilisant principalement des méthodes comportementales afin d'étudier l'apprentissage statistique dans les domaines du langage et de la musique. Nous passons en revue des recherches montrant la segmentation de séquences et l'acquisition de structures syntaxiques ou grammaticales dans des systèmes artificiels, suivies d'extensions à des matériels plus réels. La dernière section est consacrée à un bref aperçu des mécanismes potentiels sous-tendant cet apprentissage, proposés par différentes perspectives théoriques.

Mots-clés : apprentissage statistique implicite, langage, musique

One fundamental ability of the human cognitive system is to become sensitive to the regularities present in the environment. By mere exposure to the surrounding environment, and without intention to learn, perceivers are able to adapt to regularities in the world, whether perceptual, linguistic, musical, motor, or even social. Initially referred to as implicit learning by Reber (1967), and more recently referred to as statistical learning, or even implicit statistical learning (Perruchet & Pacton, 2006), the underlying mechanisms of this phenomenon have been studied also in link to language and music acquisition. Language and music are two highly structured systems; they convey statistical regularities that can be learned by mere exposure through implicit learning mechanisms. In the laboratory, the cognitive capacity of implicit statistical learning is investigated by exposing participants to miniature systems of different types, and participants' potential learning is measured with behavioral methods, such as explicit judgments, ratings, or response times, or with neurophysiological methods (e.g., M/EEG, fMRI). The present article mainly focuses on the former methods, but points the interested reader to some neuroimaging studies.

In the following, we first present laboratory research investigating implicit statistical learning in language acquisition, such as segmenting words from a speech stream, learning grammatical or syntactic structures in artificial systems, followed by extensions to more real-world materials. We then present with a similar structural organization the laboratory research investigating implicit statistical learning in music (or music-like) materials (i.e., segmentation, syntax, real-world materials). The final section is dedicated to a brief overview of the potential mechanisms underlying statistical learning, as proposed from different theoretical perspectives.

IMPLICIT STATISTICAL LEARNING IN LANGUAGE ACQUISITION

Word segmentation

For language acquisition, infants need to segment the perceived continuous speech stream to extract the words of the language. This capacity was first studied with miniature artificial languages (Figure 1) in the lab for both adults and infants by Saffran and collaborators. Classically, an artificial language paradigm is composed of two phases: an exposure phase in which the participants are presented with an unknown artificial language followed by a test phase. In one of their seminal studies (Saffran, Newport, et al., 1996), adult participants were presented with an auditory artificial language composed of six trisyllabic nonsense words generated from 12 syllables (e.g., *babupu*, *bupada*, *dutaba*, *pidabu*, *patubi*, and *tutibu*). The six words were repeated and concatenated by a synthesized language software, without direct repetitions and without acoustic cues that could aid word segmentation (no silences, no pitch or duration changes; e.g., *babupubupadadutabapatubibabupututibu*…). The only cue available to allow for segmenting the speech stream was statistical information bearing on transitional probabilities between the syllables used to constitute the artificial language. The transitional probability between two syllables X and Y is the probability that Y occurs given X. Imitating natural language (Harris, 1955; Saffran, Aslin, et al. 1996), the transitional probabilities in the artificial language were higher within words (i.e., from .31 to 1.00) than between words (i.e. from .10 to .20). After having been exposed to the artificial language for 21 minutes, participants were presented with pairs of items that consisted of one of the six words of the artificial language and either a nonword (i.e., a unit that contains the same syllables as the exposure sequence but presented in a different order) or a partword (e.g., a unit combining the last two syllables of a word with the first syllable of another word). Participants had to indicate which item of the pairs was one of the words of the artificial language. The results showed that participants were able to select above chance level the words of the language compared to either nonwords or partwords. The results of this pioneer study suggested that adult participants were sensitive to the statistical structure of the artificial language (here transitional probabilities between syllables) and were successful in correctly segmenting the words from the artificial language stream. This investigation of statistical learning leading to word segmentation in artificial languages has been extended also to 6-7-year-old children (Saffran et al.,1997) and to 8-month-old infants (Saffran, Aslin, et al., 1996) by the same research team. For infants, the paradigm was adapted with a classical familiarization-preference procedure. After the exposure phase, one of two types of stimuli was played (e.g., words versus nonwords) and its duration was controlled by the infants' gaze. With this procedure, infants showed longer listening times for nonwords and partwords compared to words, indicating a novelty preference. This study by Saffran, Aslin, et al. (1996) suggested that 8-month-old infants possess powerful learning mechanisms based on the extraction of the statistical structure of the to-be-learned material (see also Aslin et al., 1998). More recent research has extended this investigation of statistical learning to even younger infants, including newborns. Studies investigating newborns used electroencephalographic (EEG) recordings rather than behavioral methods. For example, Teinonen et al. (2009) recorded EEG of sleeping neonates from 0.5 to 2 days after birth, while presenting a continuous artificial language stream composed of ten trisyllabic words chained without acoustic markers (all syllables, whether within or between words, were separated by 200 ms). After exposure, the amplitude of a negative event-related potential was significantly larger for the first syllable of the words than for their third syllable, suggesting successful speech segmentation. Recently, similar findings have been reported with syllables being separated by a shorter delay (e.g., 150 ms in Bosseler et al., 2016) and no delay (Fló et al., 2022).

In addition to the "acoustically neutral" languages providing only statistical cues, a set of research used artificial languages that also manipulated various acoustic markers, like rhythmic and prosodic cues, such as vowel lengthening, coarticulation, or stress (e.g., Brent & Cartwright, 1996; Jusczyk et al., 1999), and even distal prosodic cues (Morrill et al., 2015). For example, adult participants' implicit statistical word learning was improved by final syllable lengthening (vs. initial syllable lengthening or no lengthening; Saffran, Newport, et al., 1996), suggesting the influence of participants' native language (i.e., English) as a strategy to parse the incoming stream (i.e. lengthening of the final syllable indicates word endings in English). Other perceptual cues can also help speech segmentation. In Perruchet and Tillmann (2010), participants were presented with an artificial language composed of six trisyllabic nonsense words. The ease with which three syllables were spontaneously perceived as a unit (referred to as the initial word-likeness, IWL) was manipulated for three of these six words. For an IWL+group, the three words, when heard in a continuous speech stream, were spontaneously perceived and recognized more often as words than were trisyllabic partwords. In an IWL–group, the three words were perceived less often as words than were the trisyllabic partwords (see also Perruchet et al., 2004). The biased words, as well as the three other words of the artificial language, were learned faster in the IWL+ group than in the IWL–group.

Pitting statistical cues against acoustic cues, notably speech surface cues (i.e., coarticulation or stress), revealed developmental changes (e.g., Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). While 9-month-old infants favor stress cues over statistical cues, 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. This outcome suggests that younger infants' sensitivity to statistical cues allows them to extract the words, and after having become more sensitive to stress patterns of their native language, they are biased to these stress patterns. Listeners' knowledge about stress patterns of their native language has also been shown to influence perceptual segmentation in a second language, acquired later in life (e.g., Sanders et al., 2002).

Follow-up studies have demonstrated that statistical learning can lead to word segmentation and the mapping of meanings to words. In Estes et al. (2007), 17-month infants first performed a classical artificial

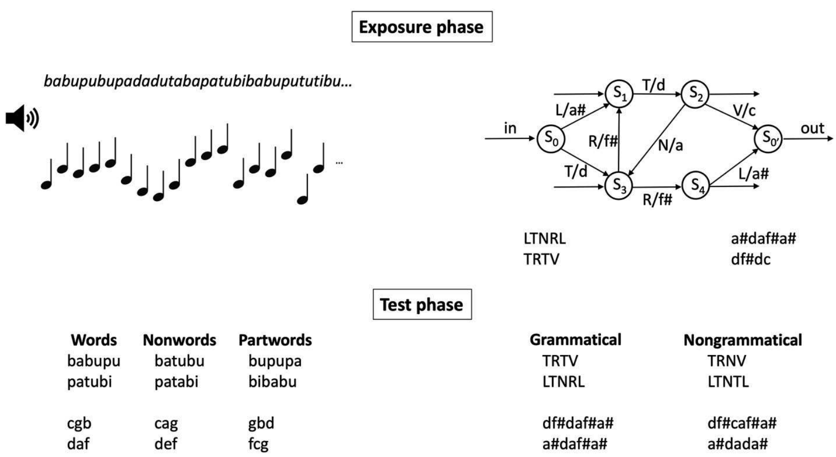

Figure 1. Illustration of the main paradigms used in implicit statistical learning with language and music materials
Note. Artificial languages being instantiated with units of three syllables or tones (left) and artificial grammars being instantiated with letters (LTNRL) or tones (a#daf#a#), played auditorily as short melodies. Examples of stimuli used in the test phases.

language protocol, immediately followed by an object-label association learning task. The infants were able to associate the object with its label, when this latter was a word of the artificial language, but not when it was a nonword or a partword. Several studies obtained similar results with natural language (e.g., Hay et al., 2011), in children (e.g., Ramos-Escobar et al., 2021), and adults (e.g., François et al., 2017; Mirman et al., 2008). Despite some differences observed as a function of age, the overall findings suggest that statistical learning can support word segmentation, emerging potential lexical candidates, and the subsequent learning of object-word association.

Syntax acquisition

In the seminal work of Reber (1967), five letters constituted the vocabulary set and the syntactic rules were defined by an artificial (finite-state) grammar (Figure 1), with which "grammatical" letter sequences were generated. A set of random letter sequences was also constituted with the same vocabulary (i.e., the same five letters). In a first experiment, an experimental group had to memorize grammatical letter sequences and a control group had to memorize random letter sequences. During the experimental session, performance continuously improved for the experimental group, while for the control group, a performance plateau was reached, and no further improvement was observed. This finding suggested that the structure of the grammatical letter sequences made the memorization more efficient. In an attempt to better understand the underlying learning processes, Reber added a testing phase after the memorization phase: participants, after having been informed about the existence of the grammar, had to make a judgment of grammaticality on new letter sequences. Some of the letter sequences were generated from the same artificial grammar as the letter sequences memorized during the exposure phase (i.e., grammatical letter sequences), whereas others did not follow the rules of the grammar (i.e., ungrammatical letter sequences). Reber observed a high proportion of correct responses, indicating that participants correctly discriminated between grammatical and ungrammatical letter sequences. Participants had become sensitive to the statistical nature of the grammatical letter sequences, although they could not verbalize on this statistical nature.

Acquiring knowledge about grammar and syntactic structures requires also to learn nonadjacent dependencies. Numerous nonadjacent dependencies are present in language syntax, for instance, between auxiliaries and inflectional morphemes (e.g., *is* work*ing*, *has* learn*ed*) or in number and gender agreement (e.g., *she* speak*s* with *her* sister). Gómez (2002) investigated the learning of such nonadjacent dependencies by using two artificial languages. The two languages (A and B) were composed of nonsense strings of three elements (e.g., *pel-wadim-jic* or *vot-kicey-rud*). For both languages, the strings began and ended with the same elements (i.e., *aXd, bXe*, and *cXf* for the first language, versus *aXe, bXf, and cXd* for the second language). The middle element was variable, as well as the size of the pool from which this middle element was drawn (from 6 to 24). The adjacent dependencies between the elements were identical between the two languages, the only difference was the nonadjacent dependencies between the first and the third elements of the strings. After having been exposed to either language A or language B, adult participants were informed that the strings they had listened to were generated based on some word-order rules, and they had to indicate whether new strings (half from language A and half from language B) followed the same order rules or not. The results showed that adult participants discriminated between strings that followed word-order rules from strings which did not, suggesting that participants learned the nonadjacent dependencies (but only when the variability of the middle element was the highest).

Using a similar material as Gómez (2002), Misyak et al. (2010) evaluated the statistical learning online with a serial reaction time (SRT) task. Six string elements were presented on 2 × 3 visuospatial grids on the screen (e.g., *pel*, *wadi*, *jic*, *vot*, *kicey*, *rud*), three of the string elements were presented auditorily in succession, and participants had to click as fast and as accurately as possible on the corresponding string written in grid cases (see Figure 2).

As in classical SRT tasks, there were several blocks, each consisting of a continuous repetition of the strings. The first six blocks were learning blocks (containing only grammatical strings), the seventh block was a block with ungrammatical strings (i.e., comparable to the random block in a classical SRT task) and the eighth block was a recovery block (i.e., presenting again grammatical strings). The results showed that the mean RT difference between the initial and the final elements of the strings gradually increased during the first six blocks, but it was drastically reduced in the ungrammatical block, and then increased again in the final grammatical block. These findings confirmed that syntax can be learned through implicit statistical learning processes. However, the online learning of nonadjacent dependencies appeared to require long exposure (i.e., no learning was evident before the fifth block).

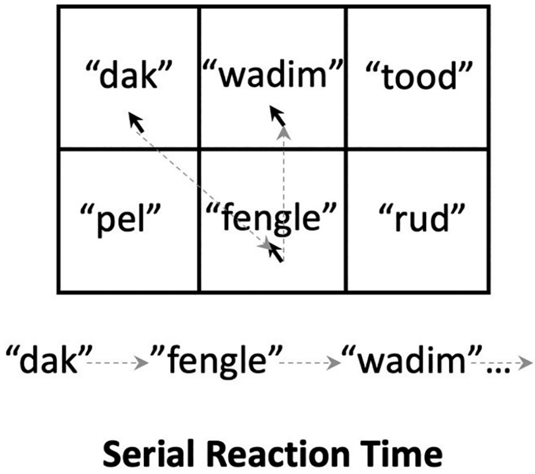

Figure 2. Illustration of a serial reaction time task used in implicit statistical learning with language material
Note. The plain arrows represent the curser of participant's computer mouse and the dotted arrows its movement as a function of the serial sequence of words written below.

The learning of new syntactic-like structures has also been shown with two other investigation methods, notably the use of 1) recursive context-free grammars, modeling some features of natural language (e.g., Rohrmeier et al., 2012), and 2) an artificial miniature language to create a set of sentences using items serving as nouns, verbs, determiners, adjectives and adverbs, thus getting closer to natural language structures (e.g., Opitz & Hofmann, 2005).

More ecological contexts

In the auditory modality, statistical learning has been studied also with a controlled set of real-world language materials (e.g., Italian, Japanese), unknown to the participants (e.g., German native speakers) to investigate whether listeners learn structural cues, including nonadjacent regularities (e.g., Müller et al., 2009). This has been investigated not only for adults, but also for infants (e.g., Friederici et al., 2011; van der Kant et al., 2020).

In the visual modality, orthographic regularities in participants' native language provide a real-world example to study implicit learning. Some of these regularities are not explicitly learned at school (i.e., double letters and their position in words), but learned based on exposure to written language. Pacton et al. (2001) evaluated whether children from grades 1 to 6, were sensitive to the fact that in French, certain letters are more frequently doubled than others (some letters being never doubled), and whether children are sensitive to the position of the double letters (i.e., only in medial position in French). Different kinds of nonwords were used, which could contain frequently double consonants (i.e., *ll*, *mm*, *ss)*, less frequently double consonants (i.e., *cc*, *dd*, *vv)*, and never occurring double letters (e.g., *uu*, *xx*, *kk*). The children had to either decide which of two nonwords is more like a word (e.g., *ommera* vs *ovvera*; *oxxubi* vs *ollubi*; *nnulor* vs *nullor*; *tillos* vs *tiilos*) or to complete nonwords (e.g., *tuba_ir*, *u_otir*) with provided patches (e.g., *LL*, *XX*). Overall, the results demonstrated that as early as grade 1, children are sensitive to the frequency of occurrence of letters and double letters, as well as to legal positions of double letters. This sensitivity was shown to increase with increasing grade levels. This study demonstrated the influence of implicit statistical learning in natural learning situations of language (i.e., natural exposure to written language). Congruent evidence has been provided by studies investigating statistical learning in the lab and reading ability (as acquired in real life), showing that better statistical learning capacity is related to better reading ability in children and adults (Arciuli & Simpson, 2012).

For school-based learning, implicit statistical learning has been assessed with new implicit learning designs. Vinter et al. (2022) compared the impact of digital games designed to teach either implicitly or explicitly the name of uppercase letters to preschool children aged from 3 to 5 years. The proposed implicit digital games promoted the automatic elicitation of associative implicit learning processes during the normal progress of play. Vinter et al. (2010) defined four main characteristics for any implicit learning situation: 1) no exposure to errors, 2) the to-be-learned material should be salient, 3) repetition is necessary, and 4) no explicit attention should be paid to the to-be-learned material. The implicit games consisted of the incidental teaching of an association between the visual image of an uppercase letter and its name (auditory stimulus) while the children were playing. The instructions given to the children and the success in the game did not require the letter-name associations. In contrast, in the explicit digital games, which were created based on educational games available on the digital game market, children's attention was explicitly drawn to the letter shape-letter name associations, mimicking a teaching procedure used at school. The children were explicitly asked to memorize the letter shape-letter name association in order to succeed in the game.

The results revealed that the implicit games were more efficient than the explicit games and the control condition (i.e., children did not play with any digital game) at ages 3 and 4. Directly comparing learning methods based on different learning mechanisms (i.e., implicit vs. explicit) at school is another way to evaluate implicit statistical learning in more ecological learning contexts.

IMPLICIT STATISTICAL LEARNING OF MUSIC

Next to the learning of language, the implicit statistical learning of music is another real-world example showing the power of the cognitive system to acquire regularities in the environment. Indeed, the native speakers becoming sensitive to the structures of the language system of their environment early in life can be compared to nonmusician listeners. Numerous research in music cognition has provided evidence that even nonmusician listeners (thus listeners without any explicit formal musical training) become sensitive to the structure of the musical system of their culture by mere exposure (i.e., tonal enculturation; e.g., Bigand & Poulin-Charronnat, 2006; Francès, 1958; Tillmann et al., 2000). For music in particular, the question emerged whether musical expertise (based on training in music schools and conservatories) might benefit to structure learning, not only of music but beyond. Divergent result patterns have been reported and methodological considerations discussed (e.g., François et al., 2012). While Rohrmeier et al. (2011) showed that musicians did not outperform nonmusicians, suggesting that the learning of a new, unfamiliar melodic system is not supported by musical expertise, others showed that learning could be boosted by musical expertise (e.g., comparing EEG responses in adult musicians and nonmusicians, see François & Schön, 2011; or in 8-year-old children having received musical or painting training, see François et al., 2013), in particular for nonadjacent regularities (Brod & Opitz, 2012). However, the potential benefit of musical training seems to be restricted to auditory material and not extending to visual statistical learning (Mandikal Vasuki et al., 2015). In addition to the benefit of musical expertise on statistical learning, a similar question can be asked for deficits, such as in congenital amusia. Some findings reported intact statistical learning capacities in amusics for both language and music-like materials (Omigie & Stewart, 2011), while others reported extended deficits, in particular when applied to tone materials (Loui &

Schlaug, 2012; Peretz et al., 2012). Note that similar investigations have been performed in the language domain, notably with the hypothesis of potential deficits in statistical learning in dyslexia or developmental language disorders, with mixed results (e.g., Bogaerts et al., 2021; Singh & Conway, 2021; West et al., 2021; for reviews and meta-analyses).

The present section follows the same structural organization as the preceding section on language, but here presenting research investigating implicit statistical learning with music (or music-like) materials and its different aspects (i.e., segmentation, syntax, real-world materials).

Music segmentation

An important debate in the statistical learning domain deals with the domain generality (i.e., a unitary learning system) versus domain specificity (i.e., a separate system for each of modality) of this learning capacity (see Saffran et al., 1999). Frost et al. (2015, 2019) proposed a theoretical framework according to which statistical learning is a domain-general process that applies to different modalities, but is subjected to specific constraints, which depend on the modality and might lead to result patterns suggesting domain-specificity. This hypothesis of a potentially domain-general learning mechanism (i.e., beyond language material) had been addressed by Saffran et al. (1999) who compared learning of artificial language systems being implemented either with syllables (as in Saffran et al., 1996) or nonlinguistic stimuli (i.e., tone sequences). An artificial language of tone words was created by replacing the syllables of the previous artificial language by tones. Six tritone "words" were created and randomly concatenated without direct repetition to generate a 21-minute-long artificial music exposure stream. After the exposure to this artificial, musical "language", nonmusician adult participants were then presented with a test phase in which pairs of tritone test items were used. In each test pair, one of the tritone items was one of the tritone "words" of the artificial music stream, while the other tritone item was either a nonword or a partword. The results revealed that, as for the syllable material, participants succeeded in selecting the words of the tone language. Supplementary analyses indicated no difference between the performance for syllable- and tone-based languages. In the same study, these results were also extended to 8-month-old infants. As for language, infants' performance was similar to adults' performance, and the tone performance was similar to the syllable performance. In Schön et al. (2008), the syllables of an artificial language were combined with musical information, thus creating sung sequences. Experiment 1 showed that when the exposure phase of a spoken artificial language, similar to that of Saffran, Newport, et al. (1996), was reduced to 7 minutes, participants were not able to discriminate between words and partwords. Experiment 2 implemented the same artificial language with sung material, with the statistical tone structure being consistent with the statistical syllable structure. Results showed that participants performed above chance level with this consistent mapping. Experiment 3 implemented the statistical structures of tones and syllables in an inconsistent way, and results revealed lower performance than in Experiment 2, albeit above-chance performance (discussed as a potential arousal effect created by the musical features). A sung advantage was also observed for newborns (François et al., 2017).

The respective influences of both acoustic and statistical features on implicit learning have been also investigated for other nonverbal material, notably musical timbre sequences (Tillmann & McAdams, 2004). The acoustic features either reinforced the statistical features, contradicted them or were neutral with respect to them. An artificial language similar to the one of Saffran, Newport, et al. (1996) was used, but syllables were replaced by musical timbres (e.g., trumpet, harp, vibraphone, clarinet, guitar). In one language, the timbres were acoustically similar within a triplet, but distant between timbre triplets, thus reinforcing the statistical structuration of the triplets. In a second language, the timbres were acoustically distant within a triplet, whereas the timbres between two different triplets (e.g., the last timbre of a triplet and the first timbre of the next triplet) were acoustically similar, thus acoustical similarity was contradicting statistical regularities. Finally, in a third language, the acoustic characteristics of timbres were neutral with regard to the statistical regularities. The results showed that participant groups being exposed to one of the three languages became all sensitive to the statistical regularities underlying the timbre sequences, and this independently of the acoustic properties of the musical timbres. All groups obtained better performance compared to control groups realizing the same test phase, but without exposure phase. However, perceptual cues can also improve statistical learning in music, notably when learning nonadjacent dependencies is evaluated. Creel et al. (2004) assessed the statistical learning of nonadjacent dependencies in tone sequences using triplets of tones (e.g., OPQ, RST, uvw, xyz) that were temporally interleaved (e.g., OuPvQwRuSvTw…). The adjacent transitional probabilities were always .05, while the nonadjacent transitional probabilities were always 1. The nonadjacent transitional probabilities being higher, they should have guided statistical learning. However, this was not the case, the participants only learned the adjacent dependencies. In two follow-up experiments, Creel et al. (2004) added a similarity cue for the nonadjacent tones (i.e., pitch frequency or timbre similarity). For instance, OPQ and RST were represented by high pitches, whereas uvw and xyz were represented by low pitches. When such prominent similarity cues are used, statistical learning was affected, and participants learned only the nonadjacent dependencies. In this case, the similarity cues could have led to auditory stream segregation and benefited to learning of event chaining in one stream. As for language, statistical learning can lead to the learning of nonadjacent dependencies in the music domain, although aided by perceptual cues.

In music, time is the second most-important form-bearing dimension beyond pitch (e.g., McAdams, 1989). Music cognition research has investigated the processing of temporal structures and regularities, including the implicit statistical learning of these regularities. The influential theory of dynamic attending by Jones (1976, 2019) hypothesized that temporal attention benefits from temporal regularities, guiding attention over cycles, thus facilitating processing of events occurring in regular structures or at expected time points. The influence of temporal regularity has been studied for implicit statistical learning for both segmentation and syntax (see below). When an artificial language was implemented only with "word units" of the same size (i.e. containing either three musical timbres or three syllables), learning was enhanced in comparison to languages implementing "word units" of different length (Hoch et al., 2013). The regularity of the units' onset, when emerging during learning, might help to guide attention to the first elements of each unit, thus reinforcing segmentation. Converging results have been reported with EEG measurements, notably with an enhanced early negativity (N1, linked to attention) emerging for the first syllable of a triplet over the exposure stream (e.g., Abla et al., 2008; Sanders et al., 2009). The advantage of rhythmic structure on learning has also been shown in newborns (Suppanen et al., 2019).

Musical syntax acquisition

As for language syntax learning, tested with different materials such as letters, syllables, or (pseudo)words, different kinds of stimuli have been used to implement artificial grammars aiming to evaluate implicit statistical learning of musical syntax. Bigand et al. (1998) used an artificial grammar similar to the one presented in Figure 1, and replaced the letters with musical timbres (i.e., gong, trumpet, piano, violin, and voice) to generate grammatical timbre sequences. During the exposure phase, participants had to memorize the timbre sequences and to indicate whether they had already heard the sequence. At the subsequent test phase, participants were informed that the timbre sequences they listened to were created by a computer program and that they were going to listen to new timbre sequences and had to indicate whether the timbre sequences were created by the same computer program (grammatical timbre sequences) or not (ungrammatical timbre sequences). Results showed that correct response rates were above chance level, and this performance was significantly better than the performance of a control group without exposure phase.

Artificial grammars have also been used by replacing the letters with tones, creating short melodies (Altmann et al., 1995). As in the other artificial grammar experiments, participants were presented with grammatical tone sequences generated by the artificial grammar. After this exposure, the test phase followed, using also a cover story with a computer program having created the melodies. Results paralleled previous findings: participants who were exposed to the grammatical tone sequences performed better in the testing phase than a control group without exposure phase.

Based on the findings that participants can acquire new, artificial grammar regularities with music-like material, the question arose whether listeners could use this newly acquired knowledge to develop online predictions (or expectations) for upcoming tones (Tillmann & Poulin-Charronnat, 2010). These predictions should thus lead to facilitated processing for expected tones in comparison to unexpected (i.e., ungrammatical) tones. In this study, the exposure phase was a classical artificial grammar exposure, with grammatical tone sequences generated from an artificial grammar (see Figure 1, right panel). However, the test phase was not the classical grammaticality judgment task, but a priming task. Participants were presented with new tone sequences, half was grammatical (i.e., generated by the same artificial grammar as the tone sequences used in the exposure phase), whereas the other half was ungrammatical, with one tone violating the artificial grammar. Participants had to listen to the tone sequences and had to make a speeded binary response on the indicated target tones for each sequence (for the purpose of this task, half of the targets were in-tune and half were out-of-tune). The response times for the tuning judgments were faster on the grammatically correct tones than the ungrammatical tones. Importantly, response times did not differ between grammatical and ungrammatical tones for participants of a control group without exposure phase. These findings suggest that participants had not only learned the regularities underlying the generation of the grammatical sequences but could also use the newly implicitly acquired knowledge to develop expectations on the tones of the sequences that facilitated the processing of expected tones.

When investigating the learning of new artificial tone systems, Rohrmeier and Cross (2013) also showed the influence of melodic, Gestalt-like features, linked to interval and contour information (e.g., listeners prefer small interval sizes, and a large interval size is followed by a small interval in the opposite direction to fill in the gap; see Narmour, 1990; Schellenberg, 1996). When interval and contour features were manipulated in an artificial structure learning context implemented with tones, the results revealed that melodic materials, containing contour and interval violations of these Gestalt-like melodic principles, impede implicit learning of the artificial tone grammar. This result is thus a case for musical material showing that implicit structure learning can be affected by prior knowledge of materials or prior preferences.

Learning of artificial-grammar structures can also be extended to more complex acoustic implementations, such as chords (more than two tones played simultaneously) creating harmonic structures (Jonaitis & Saffran 2009) or to an unfamiliar musical scale. In Loui et al. (2010), melodic sequences were generated from two artificial grammars based on an unfamiliar Bohlen-Pierce scale. Participants were exposed for 25 minutes to melodies from one of the two grammars. In the test phase, with a two-alternative forced-choice task, they had to indicate which of two melodies was generated by the same grammar as the melodies they were exposed to. The results showed that both musician and nonmusician listeners performed above chance level and successfully discriminated between the grammatical and the ungrammatical melodies. These results indicated that the musical syntax can be acquired implicitly by mere exposure to music even with unfamiliar tone elements. More recently, Loui (2022) extended these findings and provided evidence that low-level acoustical features can influence the acquired sensitivity to statistical structure in music-like material. Varying the spectral information and its link to the statistical structure allowed revealing that when the timbres included spectral information that was congruent with the to-be-learned musical scale structure, learning was the best.

Artificial grammar regularities can also be learned when implemented on the time dimension, that is with tones of different durations (Prince et al., 2018). This is in agreement with SRT studies showing learning of sequence regularities on the time dimension, with temporal patterns defined by event onsets (e.g., Brandon et al., 2012; Schultz et al., 2013). For artificial grammars of temporal regularities, learning has been shown to be restricted to the implementation with all tones being at the same pitch (isotonous presentation) and does not hold when random tones (i.e., differing in pitch) were used to install the durations. However, the same artificial grammar implemented on the pitch dimension can be learned both in an isochronous implementation and with random tone variations (Prince et al., 2018).

Learning of an artificial grammar of tones can also benefit from temporal regularities, as predicted by the dynamic attending theory (e.g., Jones, 1976). Presenting the tone sequences with a rhythmic pattern that implies a strong metrical structure leads to better learning (e.g., increased benefit of grammaticality on reaction times) than when presented with temporally irregular structures (Selchenkova, Jones et al., 2014), and even when compared to an isochronous presentation, revealing the advantage of metric binding (Jones, 2016; Selchenkova, François et al., 2014).

More ecological contexts

Music cognition research suggests that mere exposure to Western musical pieces suffices to develop implicit knowledge of the Western tonal system. Just by listening to music in everyday life, listeners become sensitive to the regularities and structures of the tonal system without being necessarily able to verbalize them (e.g., Bigand & Poulin-Charronnat, 2006; Tillmann et al., 2000). This acquisition is based on the cognitive capacity of implicit learning and has been referred to as "tonal enculturation". Francès (1958) was one of the first underlining the importance of statistical regularities in music for tonal enculturation, followed by Krumhansl (1990), Huron (2001) and others. Regularities between musical events also exist in other musical systems (e.g., Indian or Arabic music), and cultural learning and familiarity to these systems lead to auditory experiences different from those of naive listeners. Supporting evidence comes from perception studies comparing enculturated listeners and naive listeners, for example, for Balinese music (Kessler et al., 1984), Indian music (Castellano et al., 1984), Arabic music (Ayari & McAdams,

2003) or even Finnish spiritual folk hymns and North Sami yoiks (Krumhansl et al., 2000). For example, both enculturated and naive listeners show sensitivity to the sensory information present in the context, but only the enculturated listeners show the perception of fine-graded musical features that are independent of the tones or sensory information presented in the context.

In contrast to the complex real-world systems learned in everyday life, implicit statistical learning studies use reduced, simple artificial musical systems with a shorter exposure phase (see above). This allows for investigating the strengths and limits of the cognitive capacity of implicit learning in the case of nonverbal musical materials. To bridge the gap between the real-world enculturation process and the simplified experimental situation, some studies have started to use more complex systems for exposure, which are based on features of real-world music. For example, the modal melodic features of North Indian classical music (Rohrmeier & Widdess, 2017) or twelve-tone atonal music (Bigand et al., 2003). Rohrmeier and Widdess (2017) showed that Western participants unfamiliar with Indian music learned distinctive features of each mode (raga) based on exposure to melodies following one raga and succeeded in distinguishing them from melodies based on another raga. Crossing the use of one raga as either exposure or test material ensured that performance in the test phase was based on the manipulated raga and not on other musical features. Bigand et al. (2003) investigated the implicit learning of twelve-tone music in the laboratory. First, listeners were exposed to musical pieces composed with a specific 12-tone row, following atonal musical style. In the test phase, participants listened to new excerpts presented by pair and had to select the excerpt that "was composed by the same composer". More specifically, one excerpt was based on the same row as in the exposition phase, and the other excerpt on a different row. Participants (musicians and nonmusicians) performed above chance level in this test, but a control group, which had been exposed to excerpts based on both rows, did not differ from chance. This experiment suggests that the listeners became sensitive to the specific atonal structures in the exposure phase despite the complexity of the material. Taken together, these investigations using short exposures of realistic, ecologically valid materials provide evidence that incidental learning constitutes a powerful mechanism that plays a fundamental role in musical acquisition and enculturation.

WHAT IS LEARNED? THEORIES AND MODELS

Regularities in music and language systems can be acquired implicitly. However, there is still some debate about what is learned and via which mechanisms. Concerning music and language segmentation, Saffran and collaborators suggested that the learning mechanism involves the computation of transitional probabilities between events (e.g., Saffran, Aslin, et al. 1996; Saffran, Newport, et al., 1996). Learners would compute transitional probabilities between events and when the transitional probabilities are low, this would be perceived as a boundary and would allow the segmentation of the stream in "words". In a similar vein, Pearce and Wiggins (2006) proposed a computational model based on a statistical learning algorithm applied to music; it acquires melodic regularities from a reasonably sized corpus of music and is used to predict listeners' perception (see Rohrmeier & Rebuschat, 2012, for a review).

For stream segmentation (as in artificial languages), Perruchet and Vinter (1998) proposed a more parsimonious explanation with the PARSER model, a chunking model without computational assumptions. PARSER postulates that the stream would be first perceived as segmented with subjective boundaries. These first boundaries are the consequence of attentional mechanisms, which naturally segment the stream into small chunks of various lengths. Only a few of these provisional chunks are relevant for the stream structure. The fate of these chunks depends on their probability to be encountered later on again. The relevant chunks emerge through a selection process based on forgetting, the end-product of decay and interference. The less cohesive chunks among all the initial perceived chunks are eliminated, while the relevant chunks reoccurring more frequently, are strengthened and are used as new primitives to continue the stream segmentation. The PARSER model is thus based on plausible psychological mechanisms and does not use sophisticated statistical computations (see also Johnson & Tyler, 2010; Perruchet & Poulin-Charronnat, 2012; Perruchet et al., 2014; Yang, 2004, for studies that minimize the role of transitional probability computations in word segmentation).

Concerning artificial grammar and syntax acquisition, the interpretation of Reber (1967, 1989) was that participants learned an abstract representation of the rules of the grammar. This position has been challenged by Perruchet and Pacteau (1990), who defended a theoretical framework based on chunk formation. They showed that participants, who were exposed to grammatical bigrams (i.e. two letters) only, performed as well as participants who were exposed to the complete letter strings, suggesting that being exposed to fragments/chunks was sufficient to perform well in the test phase. Implicit learning performance might thus be based on the learning of chunks of letters that were partly formed based on their frequency of occurrence. This interpretation is more parsimonious as it avoids the need of an unconscious and abstract structured representation of the grammar. Several models based on chunking mechanisms have been developed both for language material (e.g., *Competitive chunking*, Servan-Schreiber & Anderson, 1990; PARSER, Perruchet & Vinter, 1998; TRACX, French et al., 2011 and TRACX2, Mareschal & French, 2017) and music material (TRACX2, Defays et al., 2023). Today, the interpretation of statistical learning based on chunk mechanisms and other interpretations based on statistical computations and rule learning are further discussed and tested. This debate continues not only for children and adults with typical development, but also for populations with atypical development or disorders (e.g., dyslexia, developmental language disorders, Parkinson disease) and across life spans (from newborns to the elderly).

Although most of the studies summarized in the present review are behavioral studies, the capacity of statistical learning has also been investigated with neuroscience methods (e.g., see Daltrozzo & Conway, 2014; Daikoku 2018; Batterink et al., 2019 for reviews), contributing to disentangle the underlying mechanisms involved in statistical learning, as well as, providing new information about the generality/specificity of the processes at work (e.g., Pesnot Lerousseau & Schön, 2021).

REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952-964. https://doi.org/10.1162/jocn.2008.20058

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 899-912. https://doi.org/10.1037/0278-7393.21.4.899

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive science*, *36*(2), 286-304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321-324. https://doi.org/10.1111/1467-9280.00063

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (Taqsīm). *Music Perception*, *21*(2), 159-216. https://doi.org/10.1525/mp.2003.21.2.159

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, *11*(3), 482-503. https://doi.org/10.1111/tops.12420

Bigand, E., D'Adamo, D. A., & Poulin, B. (2003). The implicit learning of twelve-tone music. Paper presented at the ESCOP, Grenade, Espagne.

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Current Psychology of Cognition*, *17*(3), 577-600.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130. https://doi.org/10.1016/j.cognition.2005.11.007

Bogaerts, L., Siegelman, N., & Frost, R. (2021). Statistical learning and language impairments: Toward more precise theoretical accounts. *Perspectives on Psychological Science*, *16*(2), 319-337. https://doi.org/10.1177/1745691620953082

Bosseler, A. N., Teinonen, T., Tervaniemi, M., & Huotilainen, M. (2016). Infant directed speech enhances statistical learning in newborn infants: An ERP study. *PLoS ONE*, *11*(9), Article e0162177. https://doi.org/10.1371/journal.pone.0162177

Brandon, M., Terry, J., Stevens, C. J., & Tillmann, B. (2012). Incidental learning of temporal structures conforming to a metrical framework. *Frontiers in Psychology*, *3*, Article 294. https://doi.org/10.3389/fpsyg.2012.00294

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1-2), 93-125. https://doi.org/10.1016/S0010-0277(96)00719-6

Brod, G., & Opitz, B. (2012). Does it really matter? Separating the effects of musical training on syntax acquisition. *Frontiers in Psychology*, *3*, 543. https://doi.org/10.3389/fpsyg.2012.00543

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*(3), 394-412. https://doi.org/10.1037/0096-3445.113.3.394

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119-1130. https://doi.org/10.1037/0278-7393.30.5.1119

Daikoku, T. (2018). Neurophysiological markers of statistical learning in music and language: Hierarchy, entropy, and uncertainty. *Brain Scienc*es, *8*(6), 114. https://doi.org/10.3390/brainsci8060114

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*, 437. https://doi.org/10.3389/fnhum.2014.00437

Defays, D., French, R. M., & Tillmann, B. (2023). A recurrent connectionist model of melody perception: An exploration using TRACX2. *Cognitive Science*, *47*(4), e13283. https://doi.org/10.1111/cogs.13283

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254-260.

https://doi.org/10.1111/j.1467-9280.2007.01885.x

Fló, A., Benjamin, L., Palu, M., & Deheane-Lambertz, G. (2022) Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, *12*, 4391. https://doi.org/10.1038/s41598-022-08411-w

Francès, R. (1958). *La Perception de la musique*. Paris : Librairie Philosophique J. Vrin.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038-2043. https://doi.org/10.1093/cercor/bhs180

François, C., Cunillera, T., Garcia, E., Laine, M., & Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia*, *98*, 56-67. https://doi.org/10.1016/j.neuropsychologia.2016.10.006

François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, *21*(10), 2357-2365. https://doi.org/10.1093/cercor/bhr022

François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., & Rodriguez-Fornells, A. (2017). Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. *Scientific Reports*, *7*(1), 12451. https://doi.org/10.1038/s41598-017-12798-2

François, C., Tillmann, B., & Schön, D. (2012). Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Annals of the New York Academy of Sciences*, *1252*, 108-115. https://doi.org/10.1111/j.1749-6632.2011.06395.x

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614-636. https://doi.org/10.1037/a0025255

Friederici, A. D., Müller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: Preliminary evidence from 4-month-old infants. *PLoS ONE*, *6*(3), e17920. https://doi.org/10.1371/journal.pone.0017920

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128-1153. https://doi.org/10.1037/bul0000210

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125. https://doi.org/10.1016/j.tics.2014.12.010

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436. https://doi.org/10.1111/1467-9280.00476

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*(2), 190-222. https://doi.org/10.2307/411036

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93-106. https://doi.org/10.1016/j.cogpsych.2011.06.002

Hoch, L., Tyler, M. D., & Tillmann, B. (2013). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychonomic Bulletin & Review*, *20*(1), 142-147. https://doi.org/10.3758/s13423-012-0309-8

Huron D. (2001). *Sweet anticipation. Music and the psychology of expectation*. Cambridge (Mass.): The MIT Press.

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*(4), 548-567. https://doi.org/10.1006/jmla.2000.2755

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning of word segmentation. *Developmental Science*, *13*(2), 339-345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Jonaitis, E. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science*, *33*(5), 951-968. https://doi.org/10.1111/j.1551-6709.2009.01036.x

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323-355. https://doi.org/10.1037/0033-295X.83.5.323

Jones, M. R. (2016). Musical time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (p. 125-141). Oxford: Oxford University Press.

Jones, M. R. (2019). Time will tell: A theory of dynamic attending. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190618216.001.0001

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3-4), 159-207. https://doi.org/10.1006/cogp.1999.0716

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and in the West. *Music Perception*, *2*(2), 131-165. https://doi.org/10.2307/40285289

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

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, *76*(1), 13-58. https://doi.org/10.1016/S0010-0277(00)00068-8

Loui, P. (2022). New music system reveals spectral contribution to statistical learning. *Cognition*, *224*, 105071. https://doi.org/10.1016/j.cognition.2022.105071

Loui, P., & Schlaug, G. (2012). Impaired learning of event frequencies in tone deafness. *Annals of the New York Academy of Sciences*, *1252*, 354-360. https://doi.org/10.1111/j.1749-6632.2011.06401.x

Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, *27*(5), 377-388. https://doi.org/10.1525/mp.2010.27.5.377

Mandikal Vasuki, P. R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, *342*, 112-123. https://doi.org/10.1016/j.heares.2016.10.008

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

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 181-198. https://doi.org/10.1080/07494468900640281

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271-280. https://doi.org/10.1016/j.cognition.2008.02.003

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 31. https://doi.org/10.3389/fpsyg.2010.00031

Morrill, T. H., McAuley, J. D., Dilley, L. C., Zdziarska, P. A., Jones, K. B., & Sanders, L. D. (2015). Distal prosody affects learning of novel words in an artificial language. *Psychonomic Bulletin & Review*, *22*(3), 815-823. https://doi.org/10.3758/s13423-014-0733-z

Müller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure: An ERP study in adult learners. *BMC Neuroscience*, *10*, Article 89. https://doi.org/10.1186/1471-2202-10-8

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Omigie, D., & Stewart, L. (2011). Preserved statistical learning of tonal and linguistic material in congenital amusia. *Frontiers in Psychology*, *2*, Article 109. https://doi.org/10.3389/fpsyg.2011.00109

Opitz, B., & Hofmann, J. (2015). Concurrence of rule- and similarity-based mechanisms in artificial grammar learning. *Cognitive Psychology*, *77*, 77-99. https://doi.org/10.1016/j.cogpsych.2015.02.003

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*(3), 401-426. https://doi.org/10.1037/0096-3445.130.3.401

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*(5), 377-405. https://doi.org/10.1525/mp.2006.23.5.377

Peretz, I., Saffran, J., Schön, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences*, *1252*, 361-367. https://doi.org/10.1111/j.1749-6632.2011.06429.x

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*(3), 264-275. https://doi.org/10.1037/0096-3445.119.3.264

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233-238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807-818. https://doi.org/10.1016/j.jml.2012.02.010.

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica*, *149*, 1-8. https://doi.org/10.1016/j.actpsy.2014.

Perruchet, P., & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: Human data and modeling. *Cognitive Science*, *34*(2), 255-285. https://doi.org/10.1111/j.1551-6709.2009.01074.x

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, *133*(4), 573-583. https://doi.org/10.1037/0096-3445.133.4.573

Perruchet, P., & Vinter, A. (1998). PARSER: A model of word segmentation. *Journal of Memory and Language*, *39*(2), 246-263. https://doi.org/10.1006/jmla.1998.2576

Pesnot Lerousseau, J., & Schön, D. (2021). Musical expertise is associated with improved neural statistical learning in the auditory domain. *Cerebral Cortex*, *31*(11), 4877-4890. https://doi.org/10.1093/cercor/bhab128

Prince, J. B., Stevens, C. J., Jones, M. R., & Tillmann, B. (2018). Learning of pitch and time structures in an artificial grammar setting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1201-1214. https://doi.org/10.1037/xlm0000502

Ramos-Escobar, N., Segura, E., Olivé, G., Rodriguez-Fornells, A., & François, C. (2021). Oscillatory activity and EEG phase synchrony of concurrent word segmentation and meaning-mapping in 9-year-old children. *Developmental Cognitive Neuroscience*, 51, Article 101010. https://doi.org/10.1016/j.dcn.2021.101010

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, *6*(6), 855-863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219-235. https://doi.org/10.1037/0096-3445.118.3.219

Rohrmeier, M., & Cross, I. (2013). Artificial grammar learning of melody is constrained by melodic inconsistency: Narmour's principles affect melodic learning. *PloS ONE*, *8*(7), e66174. https://doi.org/10.1371/journal.pone.0066174

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012). Implicit learning of recursive context-free grammars. *PLoS ONE*, *7*(10), Article e45885. https://doi.org/10.1371/journal.pone.0045885

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, *4*(4), 525-553. https://doi.org/10.1111/j.1756-8765.2012.01223.x

Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, *20*(2), 214-222. https://doi.org/10.1016/j.concog.2010.07.004

Rohrmeier, M., & Widdess, R. (2017). Incidental learning of melodic structure of North Indian music. *Cognitive Science*, *41*(5), 1299-1327. https://doi.org/10.1111/cogs.12404

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

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*(1), 27-52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606-621. https://doi.org/10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101-105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Sanders, L. D., Ameral, V., & Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia*, *47*(4), 1183-1186. https://doi.org/10.1016/j.neuropsychologia.2008.11.005

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech segmentation of native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research*, *45*(3), 519-530. https://doi.org/10.1044/1092-4388(2002/041)

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, *58*(1), 75-125. https://doi.org/10.1016/0010-0277(95)00665-6

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975-983. https://doi.org/10.1016/j.cognition.2007.03.005

Schultz, B. G., Stevens, C. J., Keller, P. E., & Tillmann, B. (2013). The implicit learning of metrical and nonmetrical temporal patterns. *Quarterly Journal of Experimental Psychology*, *66*(2), 360-380. https://doi.org/10.1080/17470218.2012.712146

Selchenkova, T., François, C., Schön, D., Corneyllie, A., Perrin, F., & Tillmann, B. (2014). Metrical presentation boosts implicit learning of artificial grammar. *PLoS ONE*, *9*(11), Article e112233. https://doi.org/10.1371/journal.pone.0112233

Selchenkova, T., Jones, M. R., & Tillmann, B. (2014). The influence of temporal regularities on the implicit learning of pitch structures. *Quarterly Journal of Experimental Psychology*, *67*(12), 2360-2380. https://doi.org/10.1080/17470218.2014.929155

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 592-608. https://doi.org/10.1037/0278-7393.16.4.592

Singh, S., & Conway, C. M. (2021). Unraveling the interconnections between statistical learning and dyslexia: A review of recent empirical studies. *Frontiers in Human Neuroscience*, *15*, 734179. https://doi.org/10.3389/fnhum.2021.734179

Suppanen, E., Huotilainen, M., & Ylinen, S. (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behavior & Development*, *57*, Article 101346. https://doi.org/10.1016/j.infbeh.2019.101346

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21. https://doi.org/10.1186/1471-2202-10-21

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*(4), 706-716. https://doi.org/10.1037/0012-1649.39.4.706

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*(4), 885-913. https://doi.org/10.1037/0033-295X.107.4.885

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131-1142. https://doi.org/10.1037/0278-7393.30.5.1131

Tillmann, B., & Poulin-Charronnat, B. (2010). Auditory expectations for newly acquired structures. *Quarterly Journal of Experimental Psychology*, *63*(8), 1646-1664. https://doi.org/10.1080/17470210903511228

van der Kant, A., Männel, C., Paul, M., Friederici, A. D., Höhle, B., & Wartenburger, I. (2020). Linguistic and non-linguistic non-adjacent dependency learning in early development. *Developmental Cognitive Neuroscience*, *45*, Article 100819. https://doi.org/10.1016/j.dcn.2020.100819

Vinter, A., Bard, P., Lukowski-Duplessy, H., & Poulin-Charronnat, B. (2022). A comparison of the impact of digital games eliciting explicit and implicit learning processes in preschoolers. *International Journal of Child-Computer Interaction*, *34*, 100534. https://doi.org/10.1016/j.ijcci.2022.100534

Vinter, A., Pacton, S., Witt, A., & Perruchet, P. (2010). Implicit learning, development, and education. In J.-P. Didier, & E. Bigand (Eds.), *Rethinking physical and rehabilitation medicine*. (p. 111-127). Berlin: Springer Verlag. https://doi.org/10.1007/978-2-8178-0034-9_6

West, G., Melby-Lervåg, M., & Hulme, C. (2021). Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review. *Developmental Psychology*, *57*(5), 749-770. https://doi.org/10.1037/dev0001172

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451-456. https://doi.org/10.1016/j.tics.2004.08.006.