

Review

Neurophysiological Markers of Statistical Learning in Music and Language: Hierarchy, Entropy, and Uncertainty

Tatsuya Daikoku

Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany; daikoku@cbs.mpg.de; Tel.: +81-5052157012

Received: 10 May 2018; Accepted: 18 June 2018; Published: 19 June 2018

Abstract: Statistical learning (SL) is a method of learning based on the transitional probabilities embedded in sequential phenomena such as music and language. It has been considered an implicit and domain-general mechanism that is innate in the human brain and that functions independently of intention to learn and awareness of what has been learned. SL is an interdisciplinary notion that incorporates information technology, artificial intelligence, musicology, and linguistics, as well as psychology and neuroscience. A body of recent study has suggested that SL can be reflected in neurophysiological responses based on the framework of information theory. This paper reviews a range of work on SL in adults and children that suggests overlapping and independent neural correlations in music and language, and that indicates disability of SL. Furthermore, this article discusses the relationships between the order of transitional probabilities (TPs) (i.e., hierarchy of local statistics) and entropy (i.e., global statistics) regarding SL strategies in human's brains; claims importance of information-theoretical approaches to understand domain-general, higher-order, and global SL covering both real-world music and language; and proposes promising approaches for the application of therapy and pedagogy from various perspectives of psychology, neuroscience, computational studies, musicology, and linguistics.

Keywords: statistical learning; implicit learning; domain generality; information theory; entropy; uncertainty; order; *n*-gram; Markov model; word segmentation

1. Introduction

The brain is a learning system that adapts to multiple external phenomena existing in its living environment, including various types of input such as auditory, visual, and somatosensory stimuli, and various learning domains such as music and language. By means of this wide-ranging system, humans can comprehend structured information, express their own emotions, and communicate with other people [1]. According to linguistic [2,3] and musicological studies [4,5], music and language have domain-specific structures including universal grammar, tonal pitch spaces, and hierarchical tension. Neurophysiological studies likewise suggest that there are specific neural bases for language [6,7] and music comprehension [8,9]. Nevertheless, a body of research suggests that the brain also possesses a domain-general learning system, called statistical learning (SL), that is partially shared by music and language [10,11]. SL is a process by which the brain automatically calculates the transitional probabilities (TPs) of sequential phenomena such as music and language, grasps information dynamics without an intention to learn or awareness of what we know [12,13], and further continually updates the acquired statistical knowledge to adapt to the variable

phenomena in our living environments [14]. Some researchers also indicate that the sensitivity to statistical regularities in sequences could be a by-product of chunking [15].

The SL phenomenon can partially be supported by a unified brain theory [16]. This theory tries to provide a unified account of action and perception, as well as learning under a free-energy principle [17,18], which views several keys of brain theories in the biological (e.g., neural Darwinism), physical (e.g., information theory), and neurophysiological (e.g., predictive coding) sciences. This suggests that several brain theories might be unified within a free-energy framework [19], although its capacity to unify different perspectives has yet to be established. This theory suggests that the brain models phenomena in its living environment as a hierarchy of dynamical systems that encode a causal chain structure in the sensorium to maintain low entropy [16], and predicts a future state based on the internalized model to minimize sensory reaction and optimize motor action. This prediction is in keeping with the theory of SL in the brain. That is, in SL theory, the brain models sequential phenomena based on TP distributions, grasps entropy in the whole sequences, and predicts a future state based on the internalized stochastic model in the framework of predictive coding [20] and information theory [21]. The SL also occurs in action sequences [22,23], suggesting that SL could contribute to optimization of motor action.

SL is considered an implicit and ubiquitous process that is innate in humans, yet not unique to humans, as it is also found in monkeys [24,25], songbirds [26,27], and rats [28]. The terms implicit learning and SL have been used interchangeably and are regarded as the same phenomenon [15]. A neurophysiological study [29] has suggested that conditional probabilities in the Western music corpus are reflected in the music-specific neural responses referred to as early right anterior negativity (ERAN) in event-related potential (ERP) [8,9]. The corpus study also found statistical universals in music structures across cultures [30,31]. These findings also suggest that musical knowledge may be at least partially acquired through SL. Our recent studies have also demonstrated that the brain codes the statistics of auditory sequences as relative information, such as relative distribution of pitch and formant frequencies, and that this information can be used in the comprehension of other sequential structures [10,32]. This suggests that the brain does not have to code and accumulate all received information, and thus saves some memory capacity [33]. Thus, from the perspective of information theory [21], the brain's SL is systematically efficient.

As a result of the implicit nature of SL, however, humans cannot verbalize exactly what they statistically learn. Nonetheless, a body of evidence indicates that neurophysiological and behavioural responses can unveil musical and linguistic SL effects [14,32,34–44] in the framework of predictive coding [20]. Furthermore, recent studies have detected the effects of musical training on linguistic SL of words [41,43,45–47] and the interactions between musical and linguistic SL [10] and between auditory and visual SL [44,48–50]. On the other hand, some studies have also suggested that SL is impaired in humans with domain-specific disorders such as dyslexia [51–53] and amusia [54,55], disorders that affect linguistic and music processing, respectively (though Omigie and Stewart (2011) [56] have suggested that SL is intact in congenital amusia). Thiessen et al. [57] suggested that a complete-understanding statistical learning must incorporate two interdependent processes: one is the extracting process that computes TPs (i.e., local statistics) and extracts each item, such as word segmentation, and the other one is the integration process that computes distributional information (i.e., summary statistics) and integrates information across the extracted items. The entropy and uncertainty (i.e., summary statistics), as well as TPs, are used to understand the general predictability of sequences in domain-general SL that could cover music and language in the interdisciplinary realms of neuroscience, behavioral science, modeling, mathematics, and artificial intelligence. Recent studies have suggested that SL strategies in the brain depend on the hierarchy, order [14,35,58,59], entropy, and uncertainty in statistical structures [60]. Hasson et al. [61] also indicated that certain regions or networks perform specific computations of global or summary statistics (i.e., entropy), which are independent of local statistics (i.e., TP). Furthermore, neurophysiological studies suggested that sequences with higher entropy were learned based on higher-order TP, whereas those with lower entropy were learned based on lower-order TP [59]. Thus, it is considered that information-theoretical and neurophysiological concepts on SL link each

other [62,63]. The integrated approach of neurophysiology and informatics based on the notion of order of TP and entropy can shed light on linking concepts of SL among a broad range of disciplines. Although there have been a number of studies on SL in music and language, few studies have examined the relationships between the “order” of TPs (i.e., the order of local statistics) and entropy (i.e., summary statistics) in SL. This article focuses on three themes in SL from the viewpoint of information theory, as well as neuroscience: (1) a mathematical interpretation of SL that can cover music and language and the experimental paradigms that have been used to verify SL; (2) the neural basis underlying SL in adults and children; and (3) the applicability of therapy and pedagogy for humans with learning disabilities and healthy humans.

2. Mathematical Interpretation of Brain SL Process Shared by Music and Language

2.1. Local Statistics: Nth-Order Transitional Probability

According to SL theory, the brain automatically computes TP distributions in sequential phenomena (local statistics) [35], grasps uncertainty/entropy in the whole sequences (global statistics) [61], and predicts a future state based on the internalized statistical model to minimize sensory reaction [16,20]. The TP is a conditional probability of an event B given that the latest event A has occurred, written as $P(B|A)$. The TP distributions sampled from sequential information such as music and language are often expressed by n th-order Markov models [64] or n -gram models [21] (Figure 1). Although the terminology of n -gram models has frequently been used in natural language processing, it has also recently been used in music models [65,66]. They have often been applied to develop artificial intelligence that gives computers learning abilities similar to those of the human brain, thus generating systems for data mining, automatic music composition [67–69], and automatic text classification in natural language processing [70,71]. The mathematical model of SL including n th-order Markov and $(n + 1)$ -gram models is the conditional probability of an event e_{n+1} , given the preceding n events based on Bayes' theorem:

$$P(e_{n+1} | e_n) = P(e_{n+1} \cap e_n) / P(e_n) \quad (1)$$

From the viewpoint of psychology, the formula can be interpreted as positing that the brain predicts a subsequent event e_{n+1} based on the preceding events e_n in a sequence. In other words, learners expect the event with the highest TP based on the latest n states, whereas they are likely to be surprised by an event with lower TP (Figure 2).

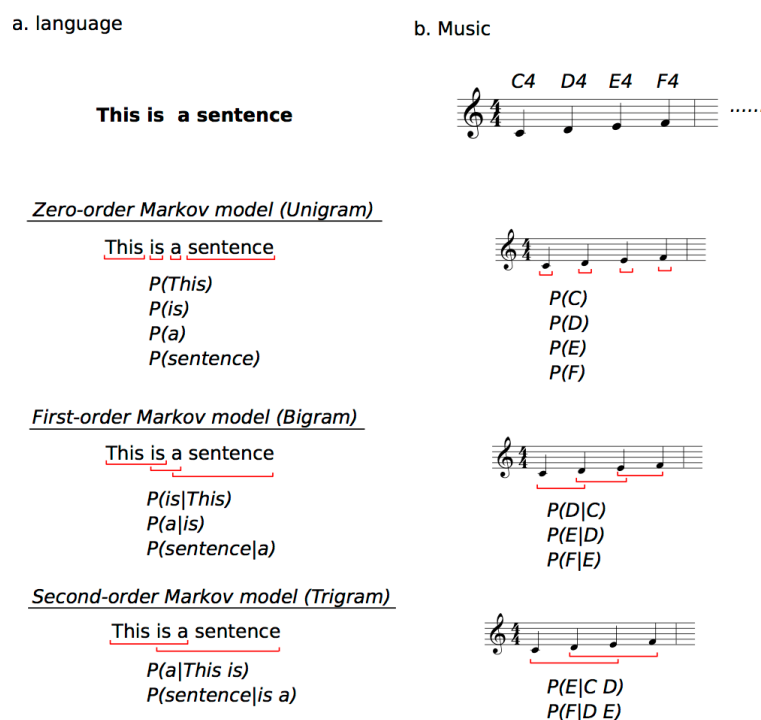


Figure 1. Example of n -gram and Markov models in statistical learning (SL) of language (a) and music (b) based on information theory. The top are examples of sequences, and the others explain how to calculate TPs ($P(e_{n+1}|e_n)$) based on zero- to second-order Markov models. They are based on the conditional probability of an event e_{n+1} , given the preceding n events based on Bayes' theorem. For instance, in language ((a), This is a sentence), the second-order Markov model represents that the "a" can be predicted based on the last subsequent two words of "This" and "is". In music ((b), C4, D4, E4, F4), second-order Markov model represents that the "E" can be predicted based on the last subsequent two tones of "C" and "D".

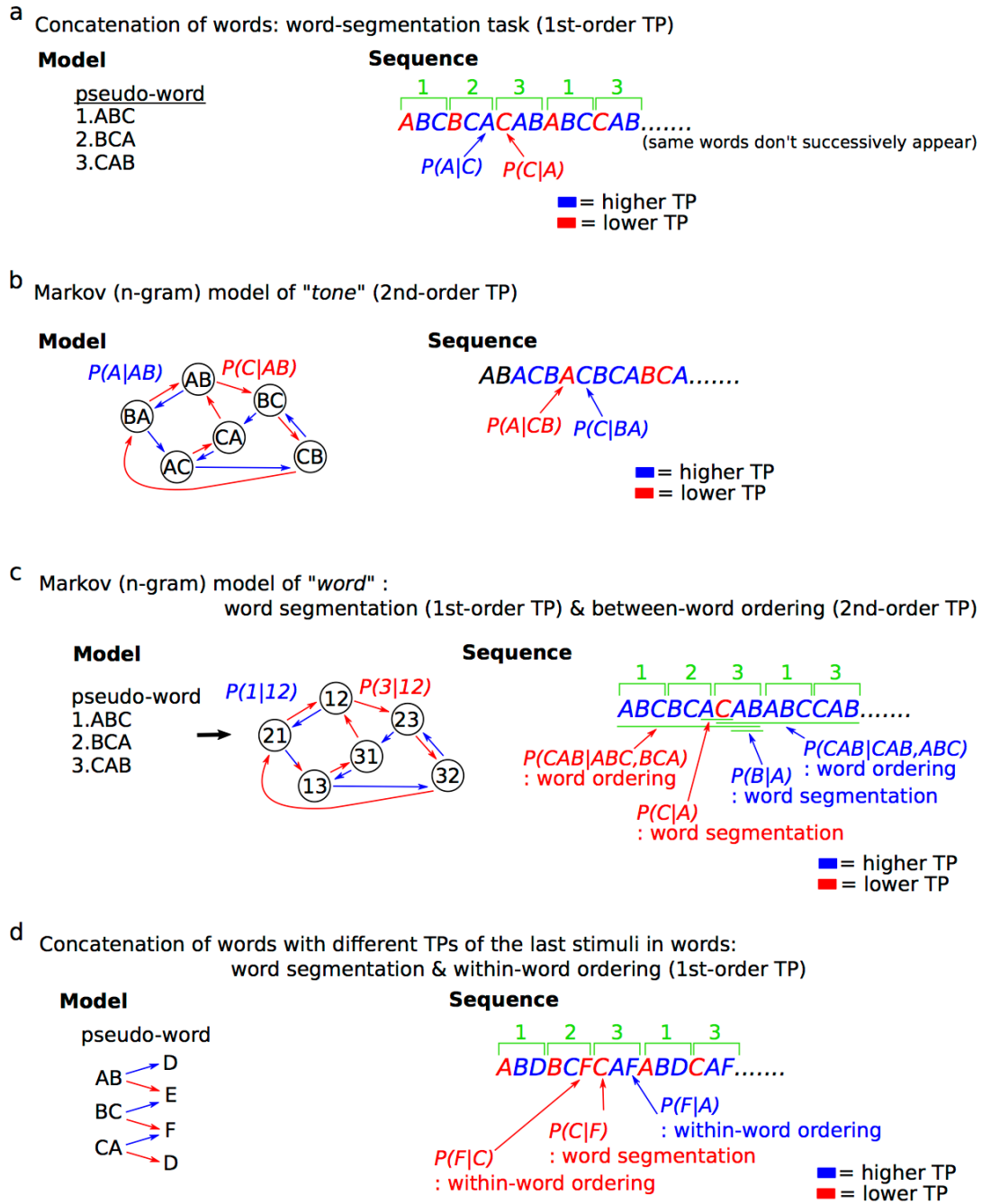


Figure 2. SL models and the sequences used in neural studies. All of the models and paradigms in sequences based on concatenation of words (a), Markov model of tone (b) and word (c), and concatenation of words with different TPs of the last stimuli in words (d) are simplified so that the characteristics of paradigms can be compared. In the example of word-segmentation paradigm (a), the same words do not successively appear. TP—transitional probability.

2.2. Global Statistics: Entropy and Uncertainty

SL models are sometimes evaluated in terms of entropy [72–75] in the framework of information theory, as done by Shannon [21]. Entropy can be calculated from probability distribution, interpreted as the average surprise (uncertainty) of outcomes [16,76], and used to evaluate the neurobiology of SL [60], as well as rule learning [77], decision making [78], anxiety, and curiosity [79,80] from the perspective of uncertainty. For instance, the conditional entropy ($H(B|A)$) in the n th order TP distribution (hereafter, Markov entropy) can be calculated from information contents:

$$H(X_{i+1}|X_i) = -\sum P(x_i) \sum P(x_{i+1}|x_i) \log_2 P(x_{i+1}|x_i) \quad (2)$$

where $H(X_{i+1}|X_i)$ is the Markov entropy; $P(X_i)$ is the probability of event x_i occurring; and $P(X_{i+1}|X_i)$ is the probability of X_{i+1} , given that X_i occurs previously. Previous articles have suggested that the degree of Markov entropy modulates human predictability in SL [61,81]. The uncertainty (i.e., global/summary statistics), as well as the TP (i.e., local statistics), of each event is applicable to and may be used to predict many types of sequential distributions, such as music and language, and to understand the predictability of a sequence (Figure 3). Indeed, entropy and uncertainty are often used to understand domain-general SL in the interdisciplinary realms of neuroscience, behavioural science, modeling, mathematics, and artificial intelligence.

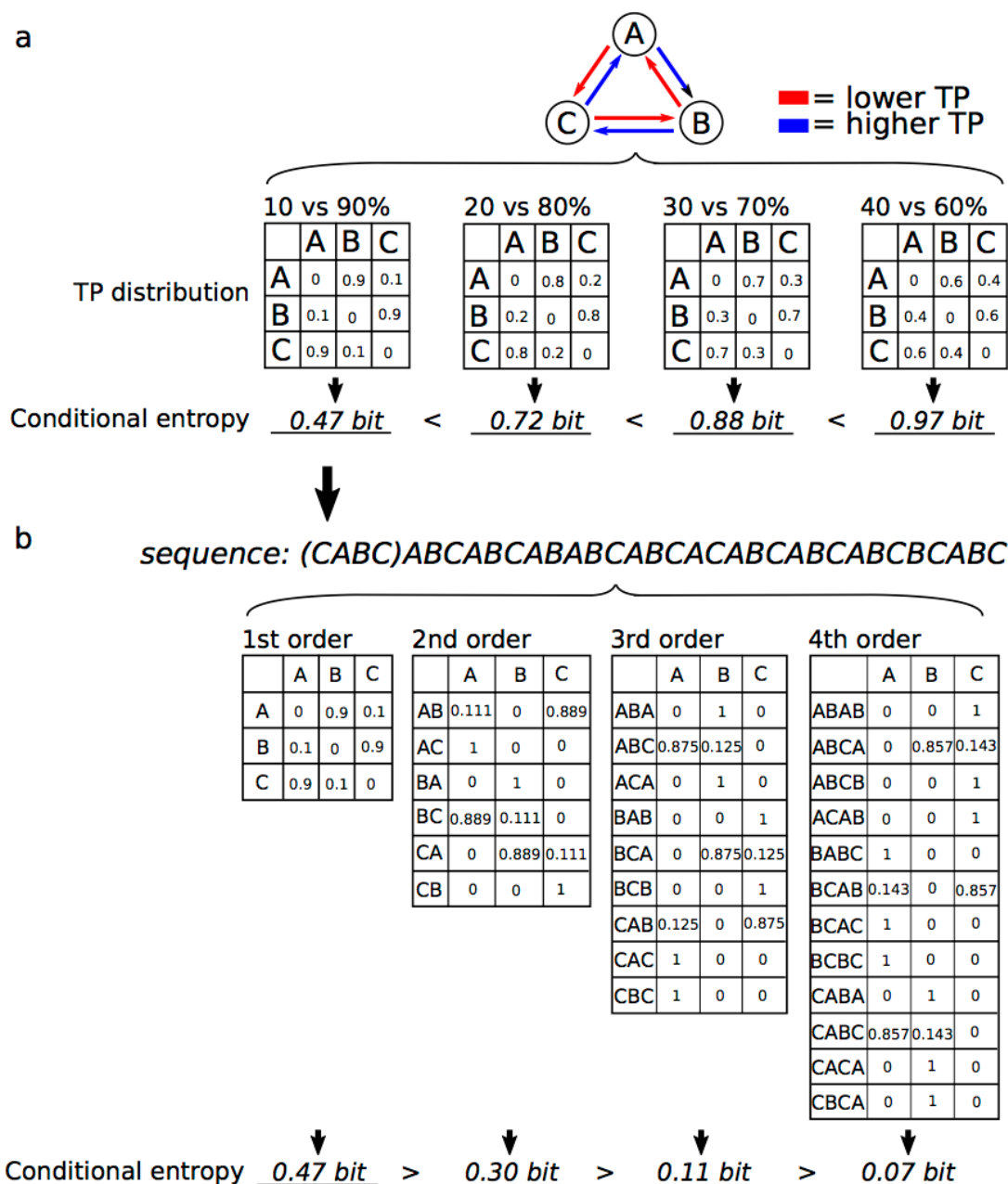


Figure 3. The entropy (uncertainty) of predictability in the framework of SL. The uncertainties depend on (a) TP ratios in a first-order Markov model (i.e., bigram model) and (b) orders of models in the TP ratio of 10% vs. 90%.

2.3. Experimental Designs of SL in Neurophysiological Studies

The word segmentation paradigm is frequently used to examine the neural basis underlying SL (e.g., [34,41,43,44,46,82–96]). This paradigm basically consists of a concatenation of pseudo-words (Figure 2a). In the pseudo-words sequence, the TP distributions based on a first-order Markov model represent lower TPs in the “first” stimulus of each word (Figure 2a: $P(B|A)$, $P(C|B)$, and $P(A|C)$) than other stimuli of word (Figure 2a: $P(C|A)$, $P(A|A)$, $P(A|B)$, $P(B|B)$, $P(B|C)$, and $P(C|C)$). When the brain statistically learns the sequences, it can identify the boundaries between words based on first-order TPs (Figure 2a) [97,98], and segment/extract each word. The SL of word segmentation based on first-order TPs has been considered as a mechanism for language acquisition in the early stages of language learning, even in infancy [12]. Recent studies have also demonstrated that SL can be performed based on within-word, as well as between-word, TPs ([40,98] for example, see Figure 2d). Although a number of studies have used a word segmentation paradigm consisting of words with a regular unit length (typically, three stimuli within a word), previous studies suggest that the unit length of words [99], the order of TPs [59], and the nonadjacent dependencies of TPs in sequences ([14,100–102] for example, see Figure 2c) can modulate the SL strategy used by the brain. Indeed, natural languages and music make use of higher-order statistics, including hierarchical, syntactical structures. To understand the brain’s higher-order SL systems in a form closer to that used for natural language and music, sequential paradigms based on higher-order Markov models have also been used in neurophysiological studies ([32,35,103] for example, see Figure 2b). Furthermore, the n th-order Markov model has been applied to develop artificial intelligence that gives computers learning and decision-making abilities similar to those of the human brain, thus generating systems for automatic music composition [67–69] and natural language processing [70,71]. Information-theoretical approaches, including information content and entropy based on n th-order Markov models, may be useful in understanding the domain-general SL, as it functions in response to real-world learning phenomena in the interdisciplinary realms of brain and computational sciences.

3. Neural Basis of Statistical Learning

3.1. Event-Related Responses and Oscillatory Activity

The ERP and event-related magnetic fields (ERF) modalities directly measure brain activity during SL and represent a more sensitive method than the observation of behavioral effects [40,41,104]. Based on predictive coding [20], when the brain encodes the TP distributions of a stimulus sequence, it expects a probable future stimulus with a high TP and inhibits the neural response to predictable external stimuli for efficiency of neural processing. Finally, the effects of SL manifest as a difference in the ERP and ERF amplitudes between stimuli with lower and higher TPs (Figure 4). Although many studies of word segmentation detected SL effects on the N400 component [43,46,88,89,93,94,105], which is generally considered to reflect a semantic meaning in language and music [106–108], auditory brainstem response (ABR) [96], P50 [41], N100 [94], mismatch negativity (MMN) [40,44,98], P200 [46,89,105], N200–250 [44,47], and P300 [83] have also been reported to reflect SL effects (Table 1). In addition, other studies using Markov models also reported that SL is reflected in the P50 [14,36,37], N100 [10,14,32,35], and P200 components [35]. Compared with later auditory responses such as N400, the auditory responses that peak earlier than 10 ms after stimulus presentation (e.g., ABR) and at 20–80 ms, which is around P50 latency, have been attributed to parallel thalamo–cortical connections or cortico–cortical connections between the primary auditory cortex and the superior temporal gyrus [109]. Thus, the suppression of an early component of auditory responses to stimuli with a higher TP in lower cortical areas can be interpreted as the transient expression of prediction error that is suppressed by predictions from higher cortical areas in a top-down connection [96]. Thus, top-down, as well as bottom-up, processing in SL may be reflected in ERP/ERF. On the other hand, SL effects on N400 have been detected in word-segmentation tasks, but not in the Markov model. TPs of a word-segmentation task are calculated based on first-order models (Figure 2a). In other words, in terms of the “order” of TP, SL

of word segmentation (i.e., sequence consisting of word concatenation) and first-order Markov model have same hierarchy of TP. Nevertheless, SL studies using the first-order Markov model did not detect learning effects of N400 (Table 1). The phenomenon of word segmentation itself has been considered as a mechanism of language acquisition in the early stages of language learning [12]. Several papers claim that the sensitivity to statistical regularities in sequences of word concatenation could be a by-product of chunking [15]. Neurophysiological effects of word segmentation, such as N400, reflecting a semantic meaning in language [106–108] may be associated with the neural basis underlying linguistic functions, as well as statistical computation itself. On the other hand, our previous study using the first-order Markov model [36] struggled to detect N400 in terms of a stimulus onset asynchrony of sequences (i.e., 500 ms). A future study will be needed to verify SL effects of N400 using the Markov model.

Table 1. Overview of neurophysiological correlations with auditory statistical learning. TP—transitional probability; ABR—auditory brainstem response; MMN—mismatch negativity; STS—superior temporal sulcus; STG—superior temporal gyrus; IFG—inferior frontal gyrus; PMC—premotor cortex; PTC—posterior temporal cortex.

Paradigms	Order of TP	Neural Correlates	References
Word segmentation	First-order	ABR	Skoe et al., 2015 [96]
		P50	Paraskevopoulos et al., 2012 [41]
		N100	Sanders et al., 2002 [94]
		MMN	Koelsch et al., 2016 [40] Moldwin et al., 2017 [98] Francois et al., 2017 [44]
		P200	De Diego Balaguer et al., 2007 [89] Francois et al., 2011 [46] Cunillera et al., 2006 [105]
		N200–250	Mandikal Vasuki et al., 2017 [47] Francois et al., 2017 [44]
		P300	Batterink et al., 2015 [83]
		N400	Cunillera et al., 2009 [88], 2006 [105] De Diego Balaguer et al., 2007 [89] Sanders et al., 2002 [94] Francois et al., 2011 [46]; 2013 [43]; 2014 [93]
		STS, STG	Farthouat et al., 2017 [91] Tremblay et al., 2012 [110] Paraskevopoulos et al., 2017 [45]
		Left IFG	Abla and Okanoya, 2008 [111] McNealy et al., 2006 [112] Paraskevopoulos et al., 2017 [45]
		PMC	Cunillera et al., 2009 [88]
		Hippocampus	Schapiro et al., 2014 [113]
Markov model	First-order	P50	Daikoku et al., 2016 [36]
		Wernicke's area	Bischoff-Grethe et al., 2000 [114]
		Hippocampus	Harrison et al., 2006 [60]
	Higher-order	P50	Daikoku et al., 2017 [14]; 2017 [37]
		N100	Furl et al., 2011 [35] Daikoku et al., 2014 [32]; 2015 [10]; 2017 [14]
		P200	Furl et al., 2011 [35]
		Right PTC	Furl et al., 2011 [35]

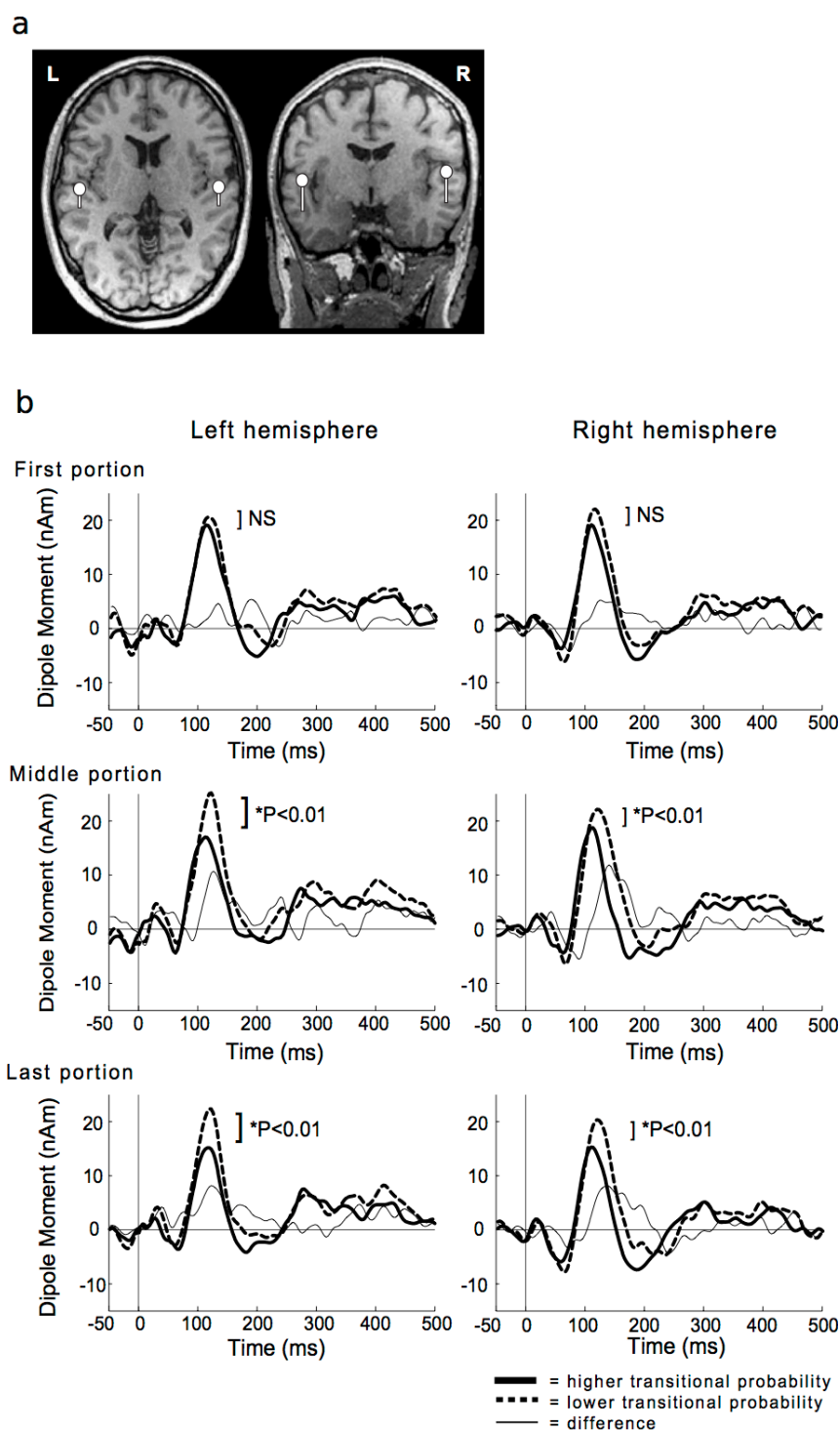


Figure 4. Representative equivalent current dipole (ECD) locations (dots) and orientations (bars) for the N100 m responses superimposed on the magnetic resonance images (a) (Daikoku et al., 2014 [32]; and the SL effects (b) (Daikoku et al., 2015 [10]) (NS = not significant). When the brain encodes the TP in a sequence, it expects a probable future stimulus with a high TP and inhibits the neural response to predictable stimuli. In the end, the SL effects manifest as a difference in amplitudes of neural responses to stimuli with lower and higher TPs (b).

It has been suggested that SL could also be reflected in oscillatory responses in the theta band [115,116]. Moreover, the human and monkey auditory cortices represent the neural marker of predictability based on SL in the form of modulations of transient theta oscillations coupling with gamma and concomitant effects [25], suggesting that SL processes are unlikely to have evolved convergently and are not unique to humans. According to previous studies, low-frequency oscillations may play an important role in speech segmentation associated with SL [73], and in tracking the envelope of the speech signal, whereas high-frequency oscillations are fundamentally involved in tracking the fine structure of speech [117]. Furthermore, there is evidence of top-down effects in low-frequency oscillations during listening to speech (up to beta band: 15–30 Hz), whereas bottom-up processing dominates in higher frequency bands [118]. Studies on the auditory oddball paradigm have also demonstrated that the power and/or coherence of theta oscillations to low-probability sounds is increased relative to high-probability sounds. Thus, many studies suggest that the lower-frequency oscillations, including theta band, are related to the prediction error [119]. Top-down predictions also control the coupling between speech and low-frequency oscillations in the left frontal areas, most likely in the speech motor cortex [120]. Although low-frequency oscillations could cover ERP components that have been suggested to reflect SL effects, the studies on oscillation and prediction imply the importance of investigating SL effects on oscillatory responses, as well as ERP.

3.2. Anatomical Mechanisms

3.2.1. Local Statistics: Transitional Probability

Neuroimaging studies have indicated that both cortical and subcortical areas play an important role in SL. For instance, the auditory association cortex, including the superior temporal sulcus (STS) [91] and superior temporal gyrus (STG) [110], contributes to auditory SL of both speech and non-speech sounds. Previous studies have also reported the effects of laterality on SL. For instance, functional magnetic resonance imaging (fMRI) [121] and near-infrared spectroscopy (NIRS) [111] studies have suggested that SL is linked to the left auditory association cortex or the left inferior frontal gyrus (IFG) [112,122], which include Wernicke's and Broca's areas, respectively. Furthermore, one previous study has indicated that brain connectivity between bilateral superior temporal sources and the left IFG is important for auditory SL [45]. On the other hand, another study has shown that the right posterior temporal cortex (PTC), which represents the high levels of the peri-Sylvian auditory hierarchy, is related to higher-order auditory SL [35] (i.e., second-order TPs). Further study will be needed to examine the relationships between the order of TPs in sequences and the neural correlations that depend on the order of TPs and hierarchy of SL.

Some studies have suggested that the sensory type of each stimulus modulates the neural basis underlying SL. For instance, some previous studies have suggested that the right hemisphere contributes to visual SL [123]. Paraskevopoulos and colleagues [50] revealed that the cortical network underlying audiovisual SL was partly common with and partly distinct from the unimodal networks of visual and auditory SL, comprising the right temporal and left inferior frontal sources, respectively. fMRI studies have also reported that Heschl's gyrus and the medial temporal lobe [124] contribute to auditory and visual SL, respectively [113], and that motor cortex activity also contributes to visual SL of action words [22]. Furthermore, Cunillera et al. [88] have suggested that the superior part of the ventral premotor cortex (PMC), as well as the posterior STG, are responsible for SL of word segmentation, suggesting that linguistic SL is related to an auditory–motor interface. Another study has suggested that the abstraction of acquired statistical knowledge is associated with a gradual shift from memory systems in the medial temporal lobe, including the hippocampus, to those of the striatum, and that this may be mediated by slow wave sleep [125].

3.2.2. Global Statistics: Entropy

Perceptive mechanisms of summary structure (i.e., global statistics) are considered to be independent of the prediction of each stimulus with different TPs (local statistics) [57,61]. Recent

studies have examined the brain systems that are responsible for encoding the uncertainty of global statistics in sequences by comparing brain activities while listening to Markov/word-concatenation and random sequences, which have lower and higher entropies, respectively. Regardless of whether music or language is assessed, the hippocampus and the lateral temporal region [88], including Wernicke's area [114], are considered to play important roles in encoding uncertainty and conditional entropy of statistical information [60]. Bischoff-Grethe et al. have also indicated that Wernicke's area may not be exclusively associated with uncertainty of language information [114]. Furthermore, uncertainty in auditory and visual statistics is coded by modality-general, as well as modality-specific, neural mechanisms [126,127], supporting the hypothesis that the neural basis underlying the brain's perception of global statistics (i.e., uncertainty), as well as local statistics (i.e., prediction of each stimulus with different TPs), is a domain-general system. Our previous neural study also suggested that reorganization of acquired statistical knowledge requires more time than the acquisition of new statistical knowledge, even if the new and previously acquired information sets have equivalent entropy levels [14]. Furthermore the results suggested that humans learn larger structures, such as phrases, first and subsequently extract smaller structures, such as words, from the learned phrases (global-to-local learning strategy). To the best of our knowledge, however, no study has yet demonstrated the differences and neural basis interactions between global and local statistics. Further study is needed to reveal how the coding of global statistics affects that of local statistics.

4. Clinical and Pedagogical Viewpoints

4.1. Disability

Although SL is a domain-general system, some studies have reported that SL is impaired in domain-specific disabilities such as dyslexia [51–53] and amusia [54,55], which are language- and music-related disabilities, respectively. Ayotte and colleagues [128] have suggested that individuals with congenital amusia fail to learn music SL but can learn linguistic SL, even if the sequences of both types have the same degree of statistical regularity [54]. Another study has suggested, in contrast, that SL is intact in amusia [56], and that individuals with amusia lack confidence in their SL ability, although they can engage in SL of music. Peretz et al. [54] stated that the input and output of the statistical computation might be domain-specific, whereas the learning mechanism might be domain-general. Furthermore, previous studies have indicated that SL ability is impaired in patients with damage to a specific area of the brain. For instance, SL is impaired in connection with hippocampal [129] and right-hemisphere damage [130]. Indeed, it has been suggested that the hippocampus plays an important role in SL [124]. One recent study indicated that auditory deprivation leads to disability of not only auditory SL [131] but also visual SL [132]. This implies that there may be specific neural mechanisms for SL that can be shared among distinct sensory modalities. Another study [133], however, suggested that a period of early deafness is not associated with SL disability. Further study is needed to clarify whether SL disability is related to temporary auditory deprivation.

4.2. Music-to-Language Transfer

4.2.1. Neural Underpinnings of SL That Overlap across Music and Language Processing

Because of the acoustic similarity [134], cortical overlap [135,136], and domain generality of SL across language and music, experienced listeners to particular spectrotemporal acoustic features, such as rhythm and pitch, in either speech or music have an advantage when perceiving similar features in the other domain [137]. According to neural studies, musical training leads to a different gray matter concentration in the auditory cortex [138] and a larger planum temporale (PT) [139–143]; the region where both language and music are processed. An ERP study has demonstrated that both the linguistic and the musical effects of SL on the N100–P200 response, which could originate in the belt and parabelt auditory regions [144,145], were larger in musicians than in non-musicians [46]. Thus, the increased PT volume associated with musical training may facilitate auditory processing

in SL. A magnetoencephalographic (MEG) study also reported that the effect of SL on the P50 response was larger in musicians than in non-musicians [41], suggesting that musical training also boosts corticofugal projections in a top-down manner regarding predictive coding [96].

Musical training could also facilitate the effects of SL on N400 [46], which is considered to be associated with IFG and PMC [88]. According to the results of a neural study, musicians have an increased gray matter density of the left IFG (i.e., Broca's area) and PMC [146]. Other studies have suggested that, during SL of word segmentation, musicians exhibit increased left-hemispheric theta coherence in the dorsal stream projecting from the posterior superior temporal (pST) and inferior parietal (IP) brain regions toward the prefrontal cortex, whereas non-musicians show stronger functional connectivity in the right hemisphere [115]. An MRI study also demonstrated that SL of word segmentation leads to pronounced left-hemisphere activity of the supratemporal plane, IP lobe, and Broca's area [147]. Thus, the left dorsal stream is considered to play an important role in SL, as well as language [7] and music learning [148].

The SL of word segmentation plays an important role in various speech abilities. Recent studies have revealed a strong link between SL of word segmentation and more general linguistic proficiency such as expressive vocabulary [149] and foreign language [150]. An fMRI study [151] has suggested that, during SL of word segmentation, participants with strong SL effects of familiar language on which they had been pretrained had decreased recruitment of fronto-subcortical and posterior parietal regions, as well as a dissociation between downstream regions and early auditory cortex, whereas participants with strong SL effects of novel language that had never been exposed showed the opposite trend. Furthermore, children with language disorders perform poorly when compared with typical developing children in tasks involving musical metrical structures [152], and have more difficulty in SL of word segmentation [153] and perception of speech rhythms [154,155]. Thus, musical training, including rhythm perception and production, is important for the development of language skills in children. Together, a body of study indicates that musical expertise may transfer to language learning [104]. It is generally considered that the left auditory cortex is more sensitive to temporal information, such as musical beat and the voice-onset (VOT) time of consonant-vowel (CV) syllables, whereas the right auditory cortex plays a role in spectral perception, such as pitch and vowel discriminations. Recent studies have indicated relationships between rhythm perception and SL [156].

Recent neural studies have demonstrated that SL of speech, pitch, timbre, and chord sequences can be performed and reflected in ERP/ERF [10,36,37,40,46]. Furthermore, the brain codes statistics of auditory sequences as relative information, such as relative distribution of pitch and formant frequencies, which could be used for comprehension of another sequential structure [10,32], suggesting that SL is ubiquitous and domain-general. On the other hand, the relative importance of acoustic features such as rhythm, pitch, intensity, and timbre varies depending on the domain, that is, music or language [157]. For instance, unlike spoken language, music contains various pitch frequencies. Recent studies have suggested that, compared with speech sequences, sung sequences with various pitches facilitate auditory SL based on word segmentation [92] and the Markov model [10]. These results further support the advantage of musical training for language SL. In addition, Hansen and colleagues have suggested that musical training also facilitates the hippocampal perception of global statistics of entropy (i.e., uncertainty) [158], as well as local statistics of each TP. Thus, musical training contributes to the improvement of SL systems in various brain regions, including the auditory cortex. Together, the facilitation of SL may be related to enhancement of the left dorsal stream via the IFG and PMC, as well as PT, enhanced low-level auditory processing in a top-down manner, and enhanced hippocampal processing. Musical training including rhythm perception contributes to these enhancements and facilitates the involvement of SL in language skills, and thus could be an important clinical and pedagogical strategy in persons with any of a variety of language-related disorders such as dyslexia [159,160] and aphasia [161].

4.2.2. Children and Adults: Critical Periods and Plasticity in the Brain

Previous studies have demonstrated that auditory SL can be performed even by sleeping neonates [85,86,162]. SL is ubiquitously performed at birth, showing that the human brain is innately prepared for it. An infant's SL extends to rhythms [163], visual stimuli [164], objects [165], social learning [23,166], and a general mechanism by which infants form meaningful representations of the environment [167]. Furthermore, infants can also learn non-adjacent statistics [101]. This suggests that SL plays an important role in an infant's syntactic learning, as well as the simple segmentation of words. These results may enable us to disentangle the respective contributions of nature and nurture in the acquisition of language and music. On the other hand, an MEG study has suggested that the strategies for language acquisition in infants could shift from domain-general SL to domain-specific processing of native language between 6 and 12 months [116], a "critical period" for language acquisition [168]. A comparable developmental change from domain-general to domain-specific learning strategies can also occur in music perception [169]. During the "critical period" of heightened plasticity, the brain is formed by sensory experience [170–172]. The development of primary cortical acoustic representations can be shaped by the higher-order TP of stimulus sequences [58]. An ERP study [173] suggested that sensitivity to speech stimuli in infants gradually shifts from accentuation to repetition during a critical period. These results may suggest that cortical reorganization depending on early experience interacts with SL [174], and that fluctuations in the degree of dependence on SL for the acquisition of language and music are part of the developmental process during critical periods. On the other hand, the SL system in the brain can be preserved even in adults (e.g., [32,35,40,41]). According to previous studies, neural plasticity can occur in adults through SL [175] and musical training [176]. In fact, there is no doubt that SL occurs in adults who are already beyond the critical periods, and that their SL ability can be modulated by auditory training. Recent studies have revealed that the process of reorganization of acquired statistical knowledge can be detected in neurophysiological responses [14]. Furthermore, a computational study on music suggested the possibility that the time-course variation of statistical knowledge over a composer's lifetime can be reflected in that composer's music from different life stages [177]. Thus, implicit updates of statistical knowledge could be enabled by the combined and interdisciplinary approach of brain, behavioral, and computational methodologies [178].

5. General Discussion

5.1. Information-Theoretical Notions for Domain-General SL: Order of TP and Entropy

SL is a domain-general and interdisciplinary notion in psychology, neuroscience, musicology, linguistics, information technology, and artificial intelligence. To generate SL models that are applicable to all of these various realms, the n th-order Markov and n -gram models based on information theory have frequently been used in natural language processing [70,71] and in the creation of automatic music composition systems [67–69]. Such models can verify hierarchies of SL based on various-order TPs. Natural languages and music include higher-order statistics, such as hierarchical syntactical structures and grammar. Thus, information-theoretical approaches, including information content and entropy based on n th-order Markov models [59,61,81], can express domain-general statistical structures closer to those of real-world language and music. The SL models are often evaluated in terms of entropy [72–75]. From a psychological viewpoint, entropy is interpreted as the average surprise (uncertainty) of outcomes [16,76]. Previous studies have demonstrated that the perception of entropy and uncertainty based on SL could be reflected in neurophysiological responses [59] and activity of the hippocampus [60]. Hasson et al. [61] indicated that certain regions or networks perform specific computations of global or summary statistics (i.e., entropy), which are independent of local statistics (i.e., TP). Furthermore, Thiessen and colleagues [57] proposed that a complete-understanding statistical learning must incorporate two interdependent processes: one is the extracting process that computes TPs and extracts each item, such as word segmentation, and the other one is the integration process that computes distributional information and integrates information across the extracted items. Our previous studies [59] investigated

correlation among entropy, order of TP, and the SL effect. As a result, the SL effects of sequences with higher entropy were lower than those with lower entropy, even when TP itself is same between these two sequences. This suggests that an evaluation of computational model of sequential information by entropy in the field of informatics may partially be able to predict learning effect in human's brain. Thus, the integrated methodology of neurophysiology and informatics based on the notion of entropy can shed light on linking the concept of SL among a broad range of disciplines. To understand the domain-general SL system that incorporates notions from both information theory and neuroscience, it is important to investigate both global and local SL.

5.2. Output of Statistical Knowledge: From Learning to Using

According to recent studies, acquired statistical knowledge contributes to the comprehension and production of complex structural information, such as music and language [179], intuitive decision-making [77,78,180–182], auditory-motor planning [183], and creativity involved in musical composition [62]. Several studies suggest that musical representation is mainly formed by a tacit knowledge [184–186]. Thus, statistical knowledge is closely tied to musical and speech expression such as composition, playing, and conversation. In addition, global statistical knowledge (i.e., entropy and uncertainty), as well as local statistical knowledge (each TP), is also supposed to contribute to decision-making [78], anxiety [80], and curiosity [79]. A number of studies have reported, however, that humans cannot verbalize exactly what they have learned statistically, even when an SL effect is detected in neurophysiological responses [14,32,34–44]. Nevertheless, our previous study suggested that statistical knowledge could alternatively be expressed via abstract medium such as musical melody [32]. In these studies, learners could behaviorally distinguish between sequences with more than eight tones with only higher TPs and those with only lower TPs, suggesting that humans can distinguish sequences with different TPs when they are provided longer sequences when compared with a conventional way in word-segmentation studies that present sequences with three tones. These studies may also suggest that that SL of auditory sequences partially interact with the Gestalt principle [5]. Furthermore, an fMRI study has suggested that the abstraction of statistical knowledge is associated with a gradual shift from the memory systems in the medial temporal lobe, including the hippocampus, to those of the striatum, and that this may be mediated by slow wave sleep [125]. Future study is needed to examine how/when statistical learning contributes to mental expression of music and language.

5.3. Applicability in Clinical and Pedagogy

Previous studies suggest that neurophysiological correlations of SL can disclose subtle individual differences that might be underestimated by behavioral levels [34,88,89,187], although recent studies showed individual differences in SL by behavioral tasks [188]. Some studies suggest that neurophysiological responses disclose SL effects, even when no SL effects cannot be detected in behavioral levels [40,41]. Neurophysiological markers of SL may at least be informative when studying less accessible populations such as infants, who are unable to deliver an obvious behavioral response [86,162]. For instance, ERP/ERF could be a useful method for the evaluation of the individual ability of SL, which is linked to individual skill in language and music learning [189,190], and which is impaired in humans with language- and music-based learning impairments such as dyslexia [51–53] and amusia [54,55]. Thus, neurophysiological markers of SL may be applicable for the evaluation of therapeutic and educational effects for patients and healthy humans [191] across any domain in which the conditional probabilities of sequential events vary systematically. Francois's findings [43] suggest the possibility of music-based remediation for children with language-based SL impairments. In addition, by using information theoretic approaches such as higher-order Markov models and entropy, SL ability can be evaluated in the form that is closest to that used in learning natural language and music [14,63]. The integration of neural, behavioral, and information-theoretical approaches may enhance our ability to evaluate SL ability in terms of both music and language.

5.4. Challenges and Future Prospects: SL in Real-World Music and Language

Although SL is generally considered domain-general, many studies also report that comprehension of language and music, which have domain-specific structures including universal grammar, tonal pitch spaces, and hierarchical tension [2–5], may rely on domain-specific neural bases [6–9,192]. Furthermore, current SL paradigms are not sufficient to account for all levels of the music- and language-learning process. Some studies suggest two steps of the learning process [193,194]. The first is SL, which shares a common mechanism among all the domains (domain generality). The second is domain-specific learning, which has different mechanisms in each domain (domain specificity). This learning process implies that, at least in an earlier step of the learning process, SL plays an essential role that covers music and language learning abilities [195]. On the other hand, few studies investigated how statistically acquired knowledge was represented in real-world communication, conversation, action, and music expression. Future studies will be needed to investigate how neural systems underlying SL contribute to comprehension and production in real-world music and language. Information-theoretical approaches based on higher-order Markov models can be used to understand SL systems in a form closer to that used for natural language and music, from a perspective of linguistics, musicology, and a unified brain theory such as the free-energy principle [16], including optimisation of action, as well as perception and learning.

6. Conclusions

This paper reviews a body of recent neural studies on SL in music and language, and discusses the possibility of therapeutic and pedagogical application. Because of a certain degree of acoustic similarity, neural overlap, and domain generality of SL between speech and music, musical training positively affects language skills in SL. Recent studies also suggested that SL strategies in the brain depend on the hierarchy, order [14,35,58,59], entropy, and uncertainty in statistical structures [60], and that certain brain regions perform specific computations of entropy that are independent of those of TP [61]. Yet few studies have investigated the relationships between the order of TPs (i.e., order of local statistics) and entropy (i.e., global statistics) in terms of SL strategies of the human brain. Information-theoretical approaches based on higher-order Markov models that can express hierarchical information dynamics as they are expressed in real-world language and music represent a possible means of understanding domain-general, higher-order, and global SL in the interdisciplinary realms of psychology, neuroscience, computational studies, musicology, and linguistics.

Funding: This work was supported by Suntory Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Ackermann, H.; Hage, S.R.; Ziegler, W. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav. Brain Sci.* **2014**, *37*, 529–604.
2. Chomsky, N. *Syntactic Structures*; Mouton: The Hague, The Netherlands, 1957.
3. Hauser, M.D.; Chomsky, N.; Fitch, W.T. The faculty of language: What is it, who has it, and how did it evolve? *Science* **2002**, *298*, 1569–1579.
4. Lerdahl, F.; Jackendoff, R. *A Generative Theory of Tonal Music*; MIT Press: Cambridge, MA, USA, 1983.
5. Jackendoff, R.; Lerdahl, F. The capacity for music: What is it, and what's special about it? *Cognition* **2006**, *100*, 33–72.
6. Friederici, A.D.; Pfeifer, E.; Hahne, A. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Brain Res. Cogn. Brain Res.* **1993**, *1*, 183–192.
7. Friederici, A.D.; Chomsky, N.; Berwick, R.C.; Moro, A.; Bolhuis, J.J. Language, mind and brain. *Nat. Hum. Behav.* **2017**, *1*, 713–722.
8. Koelsch, S.; Gunter, T.; Friederici, A.D.; Schroger, E. Brain indices of music processing: “Non-musicians” are musical. *J. Cogn. Neurosci.* **2000**, *12*, 520–541.

9. Koelsch, S. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology* **2009**, *46*, 179–190.
10. Daikoku, T.; Yatomi, Y.; Yumoto, M. Statistical learning of music- and language-like sequences and tolerance for spectral shifts. *Neurobiol. Learn. Mem.* **2015**, *118*, 8–19.
11. Saffran, J.R.; Johnson, E.K.; Aslin, R.N.; Newport, E.L. Statistical learning of tone sequences by human infants and adults. *Cognition* **1999**, *70*, 27–52.
12. Saffran, J.R.; Aslin, R.N.; Newport, E.L. Statistical learning by 8-month-old infants. *Science* **1996**, *274*, 1926–1928.
13. Cleeremans, A.; Destrebecqz, A.; Boyer, M. Implicit learning: News from the front. *Trends Cogn. Sci.* **1998**, *2*, 406–416.
14. Daikoku, T.; Yatomi, Y.; Yumoto, M. Statistical learning of an auditory sequence and reorganization of acquired knowledge: A time course of word segmentation and ordering. *Neuropsychologia* **2017**, *95*, 1–10.
15. Perruchet, P.; Pacton, S. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends Cogn. Sci.* **2006**, *10*, 233–238.
16. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138.
17. Friston, K.; Kilner, J.; Harrison, L. A free energy principle for the brain. *J. Physiol. Paris* **2006**, *100*, 70–87.
18. Friston, K.; Kiebel, S. Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2009**, *364*, 1211–1221.
19. Von Helmholtz, H. *Treatise on Physiological Optics*, 3rd ed.; Courier Corporation: Hamburg, Germany, 1909.
20. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B* **2005**, *360*, 815–836.
21. Shannon, C.E. Prediction and entropy of printed english. *Bell Syst. Tech. J.* **1951**, *30*, 50–64.
22. De Zubicaray, G.; Arciuli, J.; McMahon, K. Putting an “end” to the motor cortex representations of action words. *J. Cogn. Neurosci.* **2013**, *25*, 1957–1974.
23. Monroy, C.D.; Gerson, S.A.; Domínguez-Martínez, E.; Kaduk, K.; Hunnius, S.; Reid, V. Sensitivity to structure in action sequences: An infant event-related potential study. *Neuropsychologia* **2017**, doi:10.1016/j.neuropsychologia.2017.05.007.
24. Saffran, J.R.; Hauser, M.; Seibel, R.; Kapfhamer, J.; Tsao, F.; Cushman, F. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* **2008**, *107*, 479–500.
25. Kikuchi, Y.; Attaheri, A.; Wilson, B.; Rhone, A.E.; Nourski, K.V.; Gander, P.E.; Kovach, C.K.; Kawasaki, H.; Griffiths, T.D.; Howard, M.A., 3rd; et al. Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol.* **2017**, *15*, e2000219.
26. Lu, K.; Vicario, D.S. Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14553–14558.
27. Lu, K.; Vicario, D.S. Familiar but Unexpected: Effects of Sound Context Statistics on Auditory Responses in the Songbird Forebrain. *J. Neurosci.* **2017**, *37*, 12006–12017.
28. Toro, J.M.; Trobalón, J.B. Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* **2005**, *67*, 867–875.
29. Kim, S.G.; Kim, J.S.; Chung, C.K. The effect of conditional probability of chord progression on brain response: An MEG study. *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0017337.
30. Savage, P.E.; Brown, S.; Sakai, E.; Currie, T.E. Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8987–8992.
31. Stevens, C.J. Music perception and cognition: A review of recent cross-cultural research. *Top. Cogn. Sci.* **2012**, *4*, 653–667.
32. Daikoku, T.; Yatomi, Y.; Yumoto, M. Implicit and explicit statistical learning of tone sequences across spectral shifts. *Neuropsychologia* **2014**, *63*, 194–204.
33. Olshausen, B.A.; Field, D.J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **2004**, *14*, 481–487.
34. Abila, D.; Katahira, K.; Okanoya, K. On-line assessment of statistical learning by event related potentials. *J. Cogn. Neurosci.* **2008**, *20*, 952–964.
35. Furl, N.; Kumar, S.; Alter, K.; Durrant, S.; Shawe-Taylor, J.; Griffiths, T.D. Neural prediction of higher-order auditory sequence statistics. *Neuroimage* **2011**, *54*, 2267–2277.
36. Daikoku, T.; Yatomi, Y.; Yumoto, M. Pitch-class distribution modulates the statistical learning of atonal chord sequences. *Brain Cogn.* **2016**, *108*, 1–10.
37. Daikoku, T.; Yumoto, M. Single, but not dual, attention facilitates statistical learning of two concurrent auditory sequences. *Sci. Rep.* **2017**, *7*, 10108.

38. Daikoku, T.; Takahashi, Y.; Futagami, H.; Tarumoto, N.; Yasuda, H. Physical fitness modulates incidental but not intentional statistical learning of simultaneous auditory sequences during concurrent physical exercise. *Neurol. Res.* **2017**, *39*, 107–116.
39. Daikoku, T.; Takahashi, Y.; Tarumoto, N.; Yasuda, H. Auditory Statistical Learning during Concurrent Physical Exercise and the Tolerance for Pitch, Tempo, and Rhythm Changes. *Motor Control* **2017**, *5*, 1–24.
40. Koelsch, S.; Busch, T.; Jentschke, S.; Rohrmeier, M. Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Sci. Rep.* **2016**, *6*, 19741.
41. Paraskevopoulos, E.; Kuchenbuch, A.; Herholz, S.C.; Pantev, C. Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia* **2012**, doi:10.1016/j.neuropsychologia.2011.12.007.
42. François, C.; Tillmann, B.; Schön, D. Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 108–115.
43. François, C.; Chobert, J.; Besson, M.; Schön, D. Music training for the development of speech segmentation. *Cereb. Cortex* **2013**, *23*, 2038–2043.
44. François, C.; Cunillera, T.; Garcia, E.; Laine, M.; Rodriguez-Fornells, A. Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia* **2017**, *98*, 56–67.
45. Paraskevopoulos, E.; Chalas, N.; Bamidis, P. Functional connectivity of the cortical network supporting statistical learning in musicians and non-musicians: An MEG study. *Sci. Rep.* **2017**, *7*, 16268.
46. François, C.; Schön, D. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* **2011**, *21*, 2357–2365.
47. Mandikal Vasuki, P.R.; Sharma, M.; Ibrahim, R.; Arciuli, J. Statistical learning and auditory processing in children with music training: An ERP study. *Clin. Neurophysiol.* **2017**, *128*, 1270–1281.
48. Mitchel, A.D.; Christiansen, M.H.; Weiss, D.J. Multimodal integration in statistical learning: Evidence from the McGurk illusion. *Front. Psychol.* **2014**, *5*, 407.
49. Conway, C.M.; Christiansen, M.H. Modality-constrained statistical learning of tactile visual and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **2005**, *31*, 24–39.
50. Paraskevopoulos, E.; Chalas, N.; Kartsidis, P.; Wollbrink, A.; Bamidis, P. Statistical learning of multisensory regularities is enhanced in musicians: An MEG study. *Neuroimage* **2018**, in press.
51. Vicari, S.; Finzi, A.; Menghini, D.; Marotta, L.; Baldi, S.; Petrosini, L. Do children with developmental dyslexia have an implicit learning deficit? *J. Neurol. Neurosurg. Psychiatry* **2005**, *76*, 1392–1397.
52. Howard, J.H., Jr.; Howard, D.V.; Japikse, K.C.; Eden, G.F. Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia* **2006**, *44*, 1131–1144.
53. Menghini, D.; Hagberg, G.E.; Caltagirone, C.; Petrosini, L.; Vicari, S. Implicit learning deficits in dyslexic adults: An fMRI study. *Neuroimage* **2006**, *33*, 1218–1226.
54. Peretz, I.; Saffran, J.; Schön, D.; Gosselin, N. Statistical learning of speech, not music, in congenital amusia. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 361–367.
55. Loui, P.; Schlaug, G. Impaired learning of event frequencies in tone deafness. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 354–360.
56. Omigie, D.; Stewart, L. Preserved statistical learning of tonal and linguistic material in congenital amusia. *Front. Psychol.* **2011**, *2*, 109.
57. Thiessen, E.D.; Kronstein, A.T.; Hufnagle, D.G. The extraction and integration framework: A two-process account of statistical learning. *Psychol. Bull.* **2013**, *139*, 792–814.
58. Köver, H.; Gill, K.; Tseng, Y.T.; Bao, S. Perceptual and neuronal boundary learned from higher-order stimulus probabilities. *J. Neurosci.* **2013**, *33*, 3699–3705.
59. Daikoku, T.; Okano, T.; Yumoto, M. Relative difficulty of auditory statistical learning based on tone transition diversity modulates chunk length in the learning strategy. In Proceedings of the Biomagnetic Sendai, Japan, 22–24 May 2017; p. 75.
60. Harrison, L.M.; Duggins, A.; Friston, K.J. Encoding uncertainty in the hippocampus. *Neural Netw.* **2006**, *19*, 535–546.
61. Hasson, U. The neurobiology of uncertainty: Implications for statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 1711.
62. Pearce, M.; Wiggins, G. Auditory expectation: The information dynamics of music perception and cognition. *Top. Cogn. Sci.* **2012**, *4*, 625–652.

63. Pearce, M.T.; Ruiz, M.H.; Kapasi, S.; Wiggins, G.A.; Bhattacharya, J. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage* **2010**, *50*, 302–313.
64. Markov, A.A. *Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain*; Markov Chains; John Wiley and Sons: Hoboken, NJ, USA, 1971; Volume 1.
65. Pearce, M.T.; Wiggins, G.A. Improved methods for statistical modelling of monophonic music. *J. New Music Res.* **2004**, *33*, 367–385.
66. Rohrmeier, M.A.; Cross, I. Modelling unsupervised online-learning of artificial grammars: Linking implicit and statistical learning. *Conscious. Cogn.* **2014**, *27*, 155–167.
67. Raphael, C.; Stoddard, J. Functional harmonic analysis using probabilistic models. *Comput. Music J.* **2004**, *28*, 45–52.
68. Boenn, G.; Brain, M.; De Vos, M.; Ffitch, J. Automatic composition of melodic and harmonic music by answer set programming. In *International Conference on Logic Programming*; ICLP 2008. 5366 ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 160–174.
69. Eigenfeldt, A.; Pasquier, P. Realtime Generation of Harmonic Progressions Using Controlled Markov Selection. In *Proceedings of the ICCX-Computational Creativity Conference*, New York, NY, USA, 7–9 January 2010; pp. 16–25.
70. Brent, M.R. Speech segmentation and word discovery: A computational perspective. *Trends Cogn. Sci.* **1999**, *3*, 294–301.
71. Manning, C.D.; Schütze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.
72. Pearce, M.; Wiggins, G. Expectation in melody: The influence of context and learning. *Music Percept.* **2006**, *23*, 377–405.
73. Manzara, L.C.; Witten, I.H.; James, M. On the entropy of music: An experiment with Bach chorale melodies. *Leonardo* **1992**, *2*, 81–88.
74. Reis, B.Y. *Simulating Music Learning with Autonomous Listening Agents: Entropy, Ambiguity and Context*. Ph.D. Thesis, University of Cambridge: Cambridge, UK, 1999.
75. Cox, G. On the relationship between entropy and meaning in music: An exploration with recurrent neural networks. In *Proceedings of the Cognitive Science Society*, Portland, OR, USA, 11–14 August 2010; Volume 32.
76. Applebaum, D. *Probability and Information: An Integrated Approach*; Cambridge Univ. Press: Cambridge, UK, 2008.
77. Bach, D.R.; Dolan, R.J. Knowing how much you don't know: A neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **2012**, *13*, 572–586.
78. Summerfield, C.; de Lange, F.P. Expectation in perceptual decision making: Neural and computational mechanisms. *Nat. Rev. Neurosci.* **2014**, *15*, 745–756.
79. Loewenstein, G. The psychology of curiosity: A review and reinterpretation. *Psychol. Bull.* **1994**, *116*, 75–98.
80. Hirsh, J.B.; Mar, R.A.; Peterson, J.B. Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychol. Rev.* **2012**, *119*, 304–320.
81. Agres, K.; Abdallah, S.; Pearce, M. Information-Theoretic Properties of Auditory Sequences Dynamically Influence Expectation and Memory. *Cogn. Sci.* **2018**, *42*, 43–76.
82. Abila, D.; Okanoya, K. Visual statistical learning of shape sequences: An ERP study. *Neurosci. Res.* **2009**, *64*, 185–190.
83. Batterink, L.J.; Reber, P.J.; Neville, H.J.; Paller, K.A. Implicit and explicit contributions to statistical learning. *J. Mem. Lang.* **2015**, *83*, 62–78.
84. Batterink, L.J.; Paller, K.A. Online neural monitoring of statistical learning. *Cortex* **2017**, *90*, 31–45.
85. Bosseler, A.N.; Teinonen, T.; Tervaniemi, M.; Huotilainen, M. Infant Directed Speech Enhances Statistical Learning in Newborn Infants: An ERP Study. *PLoS ONE* **2016**, *11*, e0162177.
86. Teinonen, T.; Fellman, V.; Näätänen, R.; Alku, P.; Huotilainen, M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* **2009**, *13*, 10–21.
87. Teinonen, T.; Huotilainen, M. Implicit segmentation of a stream of syllables based on transitional probabilities: An MEG study. *J. Psycholinguist. Res.* **2012**, *41*, 71–82.

88. Cunillera, T.; Càmarà, E.; Toro, J.M.; Marco-Pallares, J.; Sebastián-Galles, N.; Ortiz, H.; Pujol, J.; Rodríguez-Fornells, A. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* **2009**, *48*, 541–553.
89. De Diego Balaguer, R.; Toro, J.M.; Rodríguez-Fornells, A.; Bachoud-Lévi, A.C. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* **2007**, *2*, e1175.
90. Buiatti, M.; Peña, M.; Dehaene-Lambertz, G. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* **2009**, *44*, 509–519.
91. Farthouat, J.; Franco, A.; Mary, A.; Delpouve, J.; Wens, V.; Op de Beeck, M.; De Tiège, X.; Peigneux, P. Auditory Magnetoencephalographic Frequency-Tagged Responses Mirror the Ongoing Segmentation Processes Underlying Statistical Learning. *Brain Topogr.* **2017**, *30*, 220–232.
92. Francois, C.; Schön, D. Learning of musical and linguistic structures: Comparing event-related potentials and behavior. *Neuroreport* **2010**, *21*, 928–932.
93. François, C.; Jaillet, F.; Takerkart, S.; Schön, D. Faster sound stream segmentation in musicians than in nonmusicians. *PLoS ONE* **2014**, *9*, e101340.
94. Sanders, L.D.; Newport, E.L.; Neville, H.J. Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* **2002**, *5*, 700–703.
95. Sanders, L.D.; Ameral, V.; Sayles, K. Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia* **2009**, *47*, 1183–1186.
96. Skoe, E.; Krizman, J.; Spitzer, E.; Kraus, N. Prior experience biases subcortical sensitivity to sound patterns. *J. Cogn. Neurosci.* **2015**, *27*, 124–140.
97. François, C.; Schön, D. Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hear. Res.* **2014**, *308*, 122–128.
98. Moldwin, T.; Schwartz, O.; Sussman, E.S. Statistical Learning of Melodic Patterns Influences the Brain's Response to Wrong Notes. *J. Cogn. Neurosci.* **2017**, *29*, 2114–2122.
99. Hoch, L.; Tyler, M.D.; Tillmann, B. Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychon. Bull. Rev.* **2013**, *20*, 142–147.
100. Frost, R.L.; Monaghan, P. Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition* **2016**, *147*, 70–74.
101. Kabdebon, C.; Pena, M.; Buiatti, M.; Dehaene-Lambertz, G. Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain Lang.* **2015**, *148*, 25–36.
102. Newport, E.L.; Aslin, R.N. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **2004**, *48*, 127–162.
103. Yumoto, M.; Daikoku, T. IV Auditory system. 5 basic function. In *Clinical Applications of Magnetoencephalography*; Tobimatsu, S., Kakigi, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 97–112.
104. Schön, D.; Francois, C. Musical expertise and statistical learning of musical and linguistic structures. *Front. Psychol.* **2011**, *2*, 167.
105. Cunillera, T.; Toro, J.M.; Sebastián-Gallés, N.; Rodríguez-Fornells, A. The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Res.* **2006**, *1123*, 168–178.
106. Koelsch, S.; Kasper, E.; Sammler, D.; Schulze, K.; Gunter, T.; Friederici, A.D. Music, language and meaning: Brain signatures of semantic processing. *Nat. Neurosci.* **2004**, *7*, 302–307.
107. Tillmann, B.; Koelsch, S.; Escoffier, N.; Bigand, E.; Lalitte, P.; Friederici, A.D.; Von Cramon, D.Y. Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *Neuroimage* **2006**, *31*, 1771–1782.
108. Kutas, M.; Federmeier, K.D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* **2011**, *62*, 621–647.
109. Adler, L.E.; Pachtman, E.; Franks, R.D.; Pecevic, M.; Waldo, M.C.; Freedman, R. Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biol. Psychiatry* **1982**, *17*, 639–654.
110. Tremblay, P.; Baroni, M.; Hasson, U. Processing of speech and non-speech sounds in the supratemporal plane: Auditory input preference does not predict sensitivity to statistical structure. *Neuroimage* **2012**, *66*, 318–332.

111. Abila, D.; Okanoya, K. Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia* **2008**, *46*, 2787–2795.
112. McNealy, K.; Mazziotta, J.C.; Dapretto, M. Cracking the language code: Neural mechanisms underlying speech parsing. *J. Neurosci.* **2006**, *26*, 7629–7639.
113. Schapiro, A.C.; Gregory, E.; Landau, B.; McCloskey, M.; Turk-Browne, N.B. The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* **2014**, *26*, 1736–1747.
114. Bischoff-Grethe, A.; Proper, S.M.; Mao, H.; Daniels, K.A.; Berns, G.S. Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *J. Neurosci.* **2000**, *20*, 1975–1981.
115. Elmer, S.; Albrecht, J.; Valizadeh, S.A.; François, C.; Rodríguez-Fornells, A. Theta Coherence Asymmetry in the Dorsal Stream of Musicians Facilitates Word Learning. *Sci. Rep.* **2018**, *8*, 4565.
116. Bosseler, A.N.; Taulu, S.; Pihko, E.; Mäkelä, J.P.; Imada, T.; Ahonen, A.; Kuhl, P.K. Theta brain rhythms index perceptual narrowing in infant speech perception. *Front. Psychol.* **2013**, *4*, 690.
117. Giraud, A.L.; Poeppel, D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.* **2012**, *15*, 511–517.
118. Fontolan, L.; Morillon, B.; Liegeois-Chauvel, C.; Giraud, A.L. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* **2014**, *5*, 4694.
119. Makeig, S. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* **1993**, *86*, 293.
120. Park, H.; Ince, R.A.; Schyns, P.G.; Thut, G.; Gross, J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **2015**, *25*, 1649–1653.
121. Asaridou, S.S.; Takashima, A.; Dediu, D.; Hagoort, P.; McQueen, J.M. Repetition Suppression in the Left Inferior Frontal Gyrus Predicts Tone Learning Performance. *Cereb. Cortex* **2016**, *26*, 2728–2742.
122. Dehaene, S.; Meyniel, F.; Wacongne, C.; Wang, L.; Pallier, C. The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron* **2015**, *88*, 2–19.
123. Roser, M.E.; Fiser, J.; Aslin, R.N.; Gazzaniga, M.S. Right hemisphere dominance in visual statistical learning. *J. Cogn. Neurosci.* **2011**, *23*, 1088–1099.
124. Reddy, L.; Poncet, M.; Self, M.W.; Peters, J.C.; Douw, L.; Van Dellen, E.; Claus, S.; Reijneveld, J.C.; Baayen, J.C.; Roelfsema, P.R. Learning of anticipatory responses in single neurons of the human medial temporal lobe. *Nat. Commun.* **2015**, *6*, 8556.
125. Durrant, S.J.; Cairney, S.A.; Lewis, P.A. Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb. Cortex* **2013**, *23*, 2467–2478.
126. Strange, B.A.; Duggins, A.; Penny, W.; Dolan, R.J.; Friston, K.J. Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Netw.* **2005**, *18*, 225–230.
127. Nastase, S.; Iacovella, V.; Hasson, U. Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum. Brain Mapp.* **2014**, *35*, 1111–1128.
128. Ayotte, J.; Peretz, I.; Hyde, K. Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* **2002**, *125*, 238–251.
129. Covington, N.V.; Brown-Schmidt, S.; Duff, M.C. The Necessity of the Hippocampus for Statistical Learning. *J. Cogn. Neurosci.* **2018**, *30*, 680–697.
130. Shaqiri, A.; Anderson, B. Priming and statistical learning in right brain damaged patients. *Neuropsychologia* **2013**, *51*, 2526–2533.
131. Studer-Eichenberger, E.; Studer-Eichenberger, F.; Koenig, T. Statistical Learning, Syllable Processing, and Speech Production in Healthy Hearing and Hearing-Impaired Preschool Children: A Mismatch Negativity Study. *Ear Hear.* **2016**, *37*, e57–e71.
132. Conway, C.M.; Pisoni, D.B.; Anaya, E.M.; Karpicke, J.; Henning, S.C. Implicit sequence learning in deaf children with cochlear implants. *Dev. Sci.* **2011**, *14*, 69–82.
133. Torkildsen, J.V.K.; Arciuli, J.; Haukedal, C.L.; Wie, O.B. Does a lack of auditory experience affect sequential learning? *Cognition* **2018**, *170*, 123–129.
134. Kraus, N.; Chandrasekaran, B. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **2010**, *11*, 599–605.
135. Schon, D.; Gordon, R.; Campagne, A.; Magne, C.; Astésano, C.; Anton, J.L.; Besson, M. Similar cerebral networks in language, music and song perception. *Neuroimage* **2010**, *51*, 450–461.
136. Peretz, I.; Vuvan, D.; Lagrois, M.E.; Armony, J.L. Neural overlap in processing music and speech. *Philos. Trans. R. Soc. B Biol. Sci.* **2015**, *370*, 68–75.

137. Ong, J.H.; Burnham, D.; Stevens, C.J.; Escudero, P. Naïve Learners Show Cross-Domain Transfer after Distributional Learning: The Case of Lexical and Musical Pitch. *Front. Psychol.* **2016**, *7*, 1189.
138. Bermudez, P.; Lerch, J.P.; Evans, A.C.; Zatorre, R.J. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* **2009**, *19*, 1583–1596.
139. Schlaug, G.; Jäncke, L.; Huang, Y.; Staiger, J.F.; Steinmetz, H. Increased corpus callosum size in musicians. *Neuropsychologia* **1995**, *33*, 1047–1055.
140. Keenan, J.P.; Thangaraj, V.; Halpern, A.R.; Schlaug, G. Absolute pitch and planum temporale. *Neuroimage* **2001**, *14*, 1402–1408.
141. Bermudez, P.; Zatorre, R.J. Differences in gray matter between musicians and nonmusicians. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 395–399.
142. Elmer, S.; Meyer, M.; Jancke, L. Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* **2012**, *22*, 650–658.
143. Elmer, S.; Hanggi, J.; Meyer, M.; Jancke, L. Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex* **2013**, *49*, 2812–2821.
144. Liegeois-Chauvel, C.; Musolino, A.; Badier, J.M.; Marquis, P.; Chauvel, P. Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* **1994**, *92*, 204–214.
145. Hackett, T.A.; Preuss, T.M.; Kaas, J.H. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* **2001**, *441*, 197–222.
146. Sluming, V.; Barrick, T.; Howard, M.; Cezayirli, E.; Mayes, A.; Roberts, N. Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage* **2002**, *17*, 1613–1622.
147. Lopez-Barroso, D.; Catani, M.; Ripolles, P.; Dell'Acqua, F.; Rodríguez-Fornells, A.; de Diego-Balaguer, R. Word learning is mediated by the left arcuate fasciculus. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13168–13173.
148. Oechslin, M.S.; Imfeld, A.; Loenneker, T.; Meyer, M.; Jancke, L. The plasticity of the superior longitudinal fasciculus as a function of musical expertise: A diffusion tensor imaging study. *Front. Hum. Neurosci.* **2010**, *3*, 76, doi:10.3389/neuro.09.076.2009.
149. Newman, R.; Ratner, N.B.; Jusczyk, A.M.; Jusczyk, P.W.; Dow, K.A. Infant's early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Dev. Psychol.* **2006**, *42*, 643–655.
150. McNealy, K.; Mazziotta, J.C.; Dapretto, M. Age and experience shape developmental changes in the neural basis of language-related learning. *Dev. Sci.* **2011**, *14*, 1261–1282.
151. Karuza, E.A.; Li, P.; Weiss, D.J.; Bulgarelli, F.; Zinszer, B.D.; Aslin, R.N. Sampling over Nonuniform Distributions: A Neural Efficiency Account of the Primacy Effect in Statistical Learning. *J. Cogn. Neurosci.* **2016**, *28*, 1484–1500.
152. Huss, M.; Verney, J.P.; Fosker, T.; Mead, N.; Goswami, U. Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex* **2011**, *47*, 674–689.
153. Evans, J.L.; Saffran, J.R.; Robe-Torres, K. Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* **2009**, *52*, 321–335.
154. Abrams, D.A.; Nicol, T.; Zecker, S.; Kraus, N. Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* **2009**, *29*, 7686–7693.
155. Goswami, U.; Wang, H.L.; Cruz, A.; Fosker, T.; Mead, N.; Huss, M. Language-universal sensory deficits in developmental dyslexia: English, Spanish and Chinese. *J. Cogn. Neurosci.* **2011**, *23*, 325–337.
156. Bouwer, F.L.; Werner, C.M.; Knetemann, M.; Honing, H. Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia* **2016**, *85*, 80–90.
157. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: Oxford, UK, 2008.
158. Hansen, N.C.; Pearce, M.T. Predictive uncertainty in auditory sequence processing. *Front. Psychol.* **2014**, *5*, 1052.
159. Habib, M.; Lardy, C.; Desiles, T.; Commeiras, C.; Chobert, J.; Besson, M. Music and dyslexia: A new musical training method to improve reading and related disorders. *Front. Psychol.* **2016**, *7*, 26.
160. Marie, C.; Magne, C.; Besson, M. Musicians and the metric structure of words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305.

161. Norton, A.; Zipse, L.; Marchina, S.; Schlaug, G. Melodic intonation therapy shared insights on how it is done and why it might help. *Neurosci. Music* **2009**, *1169*, 431–436.
162. Kudo, N.; Nonaka, Y.; Mizuno, N.; Mizuno, K.; Okanoya, K. On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* **2011**, *14*, 1100–1106.
163. Hannon, E.E.; Johnson, S.P. Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cogn. Psychol.* **2005**, *50*, 354–377.
164. Fiser, J.; Aslin, R.N. Statistical learning of higher-order temporal structure from visual shape-sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **2002**, *28*, 458–467.
165. Wu, R.; Gopnik, A.; Richardson, D.C.; Kirham, N.Z. Infants learn about objects from statistics and people. *Dev. Psychobiol.* **2011**, *47*, 1220–1229.
166. Kushnir, T.; Xu, F.; Wellman, H.M. Young children use statistical sampling to infer the preferences of other people. *Psychol. Sci.* **2010**, *21*, 1134–1140.
167. Xu, F.; Griffiths, T.L. Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition* **2011**, *120*, 299–301.
168. Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K.N.; Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **1992**, *255*, 606–608.
169. Dawson, C.; Gerken, L. From domain-general to domain-specific: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition* **2009**, *111*, 378–382.
170. Zhang, L.I.; Bao, S.; Merzenich, M.M. Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nat. Neurosci.* **2001**, *4*, 1123–1130.
171. Hensch, T.K. Critical period regulation. *Annu. Rev. Neurosci.* **2004**, *27*, 549–579.
172. Sanes, D.H.; Bao, S. Tuning up the developing auditory CNS. *Curr. Opin. Neurobiol.* **2009**, *19*, 188–199.
173. Männel, C.; Friederici, A.D. Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex* **2013**, *49*, 2788–2798.
174. Arciuli, J.; Simpson, I.C. Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Dev. Sci.* **2011**, *14*, 464–473.
175. Skoe, E.; Kraus, N. Hearing it again and again: On-line subcortical plasticity in humans. *PLoS ONE* **2010**, *5*, e13645.
176. Munte, T.F.; Altenmüller, E.; Jancke, L. The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* **2002**, *3*, 473–478.
177. Daikoku, T. Time-course variation of statistics embedded in music: Corpus study on implicit learning and knowledge. *PLoS ONE* **2018**, *13*, e0196493.
178. Arciuli, J.; Monaghan, P.; Seva, N. Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *J. Mem. Lang.* **2010**, *63*, 180–196.
179. Rohrmeier, M.; Rebuschat, P. Implicit learning and acquisition of music. *Top. Cogn. Sci.* **2012**, *4*, 525–553.
180. Berry, D.C.; Dienes, Z. *Implicit Learning: Theoretical and Empirical Issues*; Lawrence Erlbaum: Hove, UK, 1993.
181. Reber, A.S. *Implicit Learning and Tacit Knowledge. An Essay on the Cognitive Unconscious*; Oxford University Press: New York, NY, USA, 1993.
182. Perkovic, S.; Orquin, J.L. Implicit Statistical Learning in Real-World Environments Leads to Ecologically Rational Decision Making. *Psychol. Sci.* **2017**, *29*, 34–44.
183. Norgaard, M. How jazz musicians improvise: A central role of auditory and motor patterns. *Music Percept.* **2014**, *31*, 271–287.
184. Bigand, E.; Poulin-Charronnat, B. Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* **2006**, *100*, 100–130.
185. Ettlinger, M.; Margulis, E.H.; Wong, P.C.M. Implicit memory in music and language. *Front. Psychol.* **2011**, *211*, doi:10.3389/fpsyg.2011.00211.
186. Huron, D. Two challenges in cognitive musicology. *Top. Cogn. Sci.* **2012**, *4*, 678–684.
187. McLaughlin, J.; Osterhout, L.; Kim, A. Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nat. Neurosci.* **2004**, *7*, 703–704.
188. Siegelman, N.; Bogaerts, L.; Christiansen, M.H.; Frost, R. Towards a theory of individual differences in statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 1711.

189. Arciuli, J.; Simpson, I.C. Statistical learning is related to reading ability in children and adults. *Cogn. Sci.* **2012**, *36*, 286–304.
190. Kidd, E.; Arciuli, J. Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Dev.* **2016**, *87*, 184–193.
191. Shaqiri, A.; Anderson, B.; Danckert, J. Statistical learning as a tool for rehabilitation in spatial neglect. *Front. Hum. Neurosci.* **2013**, *7*, 224.
192. Daikoku, T.; Ogura, H.; Watanabe, M. The variation of hemodynamics relative to listening to consonance or dissonance during chord progression. *Neurol. Res.* **2012**, *34*, 557–563.
193. Ellis, R. Implicit and explicit learning, knowledge and instruction. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*; Ellis, R., Loewen, S., Elder, C., Erlam, R., Philip, J., Reinders, H., Eds.; Multilingual Matters: Bristol, UK, 2009; pp. 3–25.
194. Jusczyk, P.W. How infants begin to extract words from speech. *Trends Cogn. Sci.* **1999**, *3*, 323–328.
195. Archibald, L.M.; Joanisse, M.F. Domain-specific and domain-general constraints on word and sequence learning. *Mem. Cogn.* **2013**, *41*, 268–280.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).