scientific reports

Check for updates

OPEN

Temporal dynamics of statistical learning in children's song contributes to phase entrainment and production of novel information in multiple cultures

Tatsuya Daikoku

Statistical learning is thought to be linked to brain development. For example, statistical learning of language and music starts at an early age and is shown to play a significant role in acquiring the delta-band rhythm that is essential for language and music learning. However, it remains unclear how auditory cultural differences affect the statistical learning process and the resulting probabilistic and acoustic knowledge acquired through it. This study examined how children's songs are acquired through statistical learning. This study used a Hierarchical Bayesian statistical learning (HBSL) model, mimicking the statistical learning processes of the brain. Using this model, I conducted a simulation experiment to visualize the temporal dynamics of perception and production processes through statistical learning among different cultures. The model learned from a corpus of children's songs in MIDI format, which consists of English, German, Spanish, Japanese, and Korean songs as the training data. In this study, I investigated how the probability distribution of the model is transformed over 15 trials of learning in each song. Furthermore, using the probability distribution of each model over 15 trials of learning each song, new songs were probabilistically generated. The results suggested that, in learning processes, chunking and hierarchical knowledge increased gradually through 15 rounds of statistical learning for each piece of children's songs. In production processes, statistical learning led to the gradual increase of delta-band rhythm (1–3 Hz). Furthermore, by combining the acquired chunks and hierarchy through statistical learning, statistically novel music was generated gradually in comparison to the original songs (i.e. the training songs). These findings were observed consistently, in multiple cultures. The present study indicated that the statistical learning capacity of the brain, in multiple cultures, contributes to the acquisition and generation of delta-band rhythm, which is critical for acquiring language and music. It is suggested that cultural differences may not significantly modulate the statistical learning effects since statistical learning and slower rhythm processing are both essential functions in the human brain across cultures. Furthermore, statistical learning of children's songs leads to the acquisition of hierarchical knowledge and the ability to generate novel music. This study may provide a novel perspective on the developmental origins of creativity and the importance of statistical learning through early development.

Statistical learning and emergence of individuality

Music is a ubiquitous element in many cultures and is instrumental in shaping the unique characteristics that define each one. Through the learning of music, humans develop knowledge that is culturally specific or shared across cultures. Although music plays a crucial role in human development, the underlying mechanisms that enable the brain to learn and produce music remain poorly understood.

[1]Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [2]Center for Brain, Mind and KANSEI Sciences Research, Hiroshima University, Hiroshima, Japan. email: daikoku.tatsuya@mail.u-tokyo.ac.jp

In recent years, a growing body of studies has tried to explain the learning and production mechanisms of music based on the general principle of predictive processing in the brain[1]. Predictive processing of music works to minimize the prediction error (i.e. surprise) between the bottom-up auditory signals including melody, rhythm, and musical chords and the top-down predictive signals based on internal music models of the brain[2,3].

Within the framework of the brain's predictive processing, it has been proposed that the acquisition of musical knowledge may be underpinned by statistical learning mechanisms[4,5]. Statistical learning is a brain's function that is closely linked to brain development[6] and contributes to the perception and production of music and language. The basic mechanism involves calculating the statistical probability of environmental information (particularly the transition probability of sequential information) and the uncertainty of probability distribution, and predicting future information based on an internal probabilistic model acquired through statistical learning.

Importantly, uncertainty and probability are not universally inherent in music per se but are instead shaped by the amount of an individual's auditory experiences. For example, when non-native individuals listen to culturally specific ethnic music, the uncertainty in predicting the probable subsequent sound is typically high, making prediction more difficult. In contrast, people within the ethnic group who regularly listen to such music often find it easier to predict sounds (i.e. less surprise) due to their lower level of uncertainty[7]. This is a result of individuals constantly updating their internal models through extended periods of statistical learning, thereby generating an appropriate music probability model within the culture. Neural and computational studies have suggested that the individual traits related to music "production" (i.e. composition)[8–10] and perception[11–13] are associated with the statistical familiarity (or experience) acquired from a particular musical culture via statistical learning.

Reliability of prediction

Auditory experience affects not only uncertainty, prediction, and surprise but also the "reliability" of predictions. For example, the degree of reliability varies even when the transition probability is identical at 90% between the case of experiencing A-to-B transitions nine times out of 10 trials and the case of experiencing them 90 times out of 100 trials. Additionally, the brain can rely more on events with low transition probability if they have been experienced 10 out of 100 times instead of only once out of 10. Intuitively, one is in a state of recognizing that this event is "reliably" unpredictable.

Neurophysiological studies[14,15] have demonstrated a progressive representation of statistical learning effects with an increase in the number of learning repetitions. This indicates that the brain's statistical learning is based on Bayesian inference, which enhances the reliability of probabilities with experience, as opposed to maximum likelihood estimation, which does not vary with learning repetitions. However, most studies of statistical learning have referred to maximum likelihood estimation based on Markov models or n-gram models that do not consider the "reliability" of probabilities, and thus, have not considered the effect of learning repetitions.

To resolve such problems of these statistical learning studies, this study developed a computational model of the brain's statistical learning, referred to as "Hierarchical Bayesian Statistical Learning (HBSL)" incorporating the Bayesian reliability of probabilities into a Markov model. I then used this model to examine the learning process when lots of music are repetitively learned.

Hierarchically structured building and its phase entrainment

Statistical learning essentially contributes to chunking, which involves identifying information units with high transition probabilities from sequential information such as short phrases and words. Prior research has investigated the neural and computational mechanisms underlying chunking through statistical learning. In contrast, recent studies have suggested two types of "*hierarchical*" statistical learning systems[10,16]. The first system constitutes the fundamental function of statistical learning, which groups chunks of information with high transition probabilities and integrates them into a cohesive unit. The second system involves statistical learning that arranges various chunked units to form a hierarchical syntactic structure (Fig. 1). Therefore, statistical learning plays a critical role in acquiring the hierarchy, a unique and essential feature of language and music[17].

Particularly, the hierarchical structure of rhythms is essential for the acquisition of music and language[18]. The lower hierarchy corresponds to a frequency band of syllable and musical notes (e.g. quarter notes) around 4–12 Hz while the higher hierarchy corresponds to prosody, intonation, and long musical notes (e.g. half notes) around 1–3 Hz[19]. There are also rhythms around 12–30 Hz that correspond to phonemes or sound onsets at even lower levels of the hierarchy.

Such a hierarchical structure of rhythm can be visualized from the amplitude modulation (AM) envelope of sound waveforms[20–22]. Evidence has shown that neural oscillation in the human brain phase synchronizes the AM hierarchy of auditory rhythm[23]. Such a synchronization or phase entrainment contributes to the parsing of the sound signal into each unit of the hierarchy[24]. Further, even in the absence of acoustic (e.g. AM) cues to hierarchical structure, it has been demonstrated that the brain is phase entrained to the hierarchical rhythm simply from the syntactic but not acoustic structure[25]. That is, the brain possesses the capacity to interpret and phase-synchronize with rhythm based on the syntax of sound sequences, even in the absence of an acoustically rhythmic hierarchy[25,26]. Similarly, neural oscillations can synchronize with statistically chunked rhythm, transcending the need for an overt rhythmic hierarchy in the acoustic templates of sound sequences[27,28]. This suggests the brain's ability to discern hierarchical structures of chunk rhythms, regardless of the acoustic rhythmic hierarchy[29]. Further, this raises the possibility that, even if the input information has a weak acoustic hierarchy, the brain can produce output information with an acoustic hierarchical structure by syntactically processing the hierarchical structure in the brain.

A study comparing the hierarchical rhythmic structure across various types of sounds (nature, speech, instrumental music, song, animal sounds, etc.) has demonstrated that the slower band hierarchy (i.e. 1–3 Hz) is especially pronounced in children's songs and speech directed towards infants and children (referred to as infant/

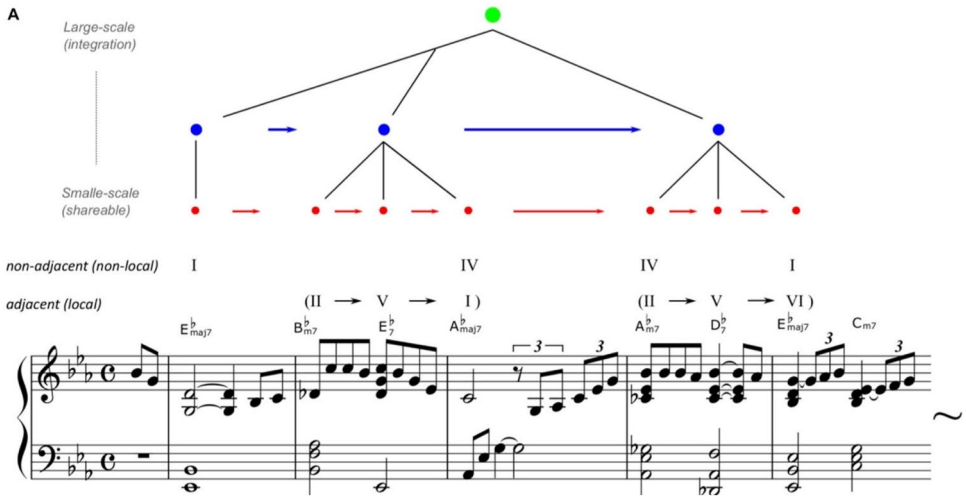

Figure 1. An example of hierarchical statistical learning of music. Reprinted from Daikoku et al.[10]. Misty by Errol Garner, composed in 1954. The arrangement, chord names, and symbols are simplified (just major/minor, flat, and 7th note) to account for the two-five-one (II–V(7)–I) progression. For example, jazz music has general regularities in chord sequences such as the so-called "two-five-one (II–V–I) progression." Such syntactic progression frequently occurs in music, and therefore, the statistics of the sequential information have high transitional probability and low uncertainty. Thus, once a person has learned the statistical characteristics, it can be chunked as a commonly used unit among improvisers. In contrast, the ways of combining the chunked units (e.g. combining among different blue circled units) are different between musicians.

child-directed speech) compared to other types of sounds[19] and in at least one other culture[30]. Further, evidence has shown that the infant is first phase-entrained to the slower rhythm (1–3 Hz) in infant-directed speech[31], suggesting that such a slower-band rhythm is important for early learning of language and music. Thus, children's songs and child-directed speech may emphasize 1–3 Hz rhythms to facilitate the acquisition of rhythm hierarchy in early auditory learning. Alternatively, it could be hypothesized that children's songs emphasize 1–3 Hz rhythms because children syntactically and acoustically learn these rhythms first.

Evidence has shown that the ability of 1–3 Hz phase entrainment is closely linked to statistical learning capacity[32]. By statistical learning, the neural oscillations can be entrained to the statistically chunked information that has a slower rhythm (e.g. 1 Hz word rhythm that has three 3 Hz syllables[27,28]). Thus, statistical learning plays a critical role in phase entrainment of the slower rhythm around 1-3 Hz, essential for early learning of music. However, it is unclear how music stimuli are chunked into a 1–3 Hz unit and how they are manifested as 1–3 Hz rhythms when producing a new musical information through statistical learning, and whether different internal models that may vary with culture and experience affect the processing of 1–3 Hz rhythms.

Purpose of the present study

The present study examined the temporal dynamics of perception and production processes through statistical learning, using the HBSL model that considers both reliability and hierarchy, mimicking the statistical learning processes of the brains. The model learned from a corpus of children's songs in MIDI format, which consists of English, German, Spanish, Japanese, and Korean songs as the training data. This study investigated how the probability distribution of the model is transformed over 15 trials of learning in each song. Furthermore, using the probability distribution of these 15 learned models in each culture, new songs were probabilistically generated.

This study hypothesized that statistical learning leads to a gradual increase of 1–3 Hz rhythms in music production. Further, since statistical learning and slower rhythm processing are both essential functions in the human brain across multiple cultures, it was hypothesized that cultural differences do not significantly modulate the statistical learning effects.

Results

Learning process

The total number of chunks and hierarchy generated during statistical learning was measured in each of the 15 trials. All results were shown in an external source (https://osf.io/cqmz8/?view_only=b95d94626a364700adb9e1e94384525d). The results revealed that over the course of 15 statistical learning trials, both the number of chunks (as demonstrated by the top left quadrant of Fig. 2) and hierarchy (as demonstrated by the right quadrant of Fig. 2) gradually increased. Further, the total Bayesian surprise (or total prediction errors) that occurred in each statistical learning trial was measured by the Kullback–Leibler divergence between a probability distribution P(x) before exposure to a stimulus and a probability distribution after exposure to a stimulus. The results showed that the Bayesian surprise showed a gradual decrease over the course of 15 statistical learning trials, as demonstrated by the data located in the bottom left quadrant of Fig. 2. Finally, these findings were consistent across the cultures included in this study.

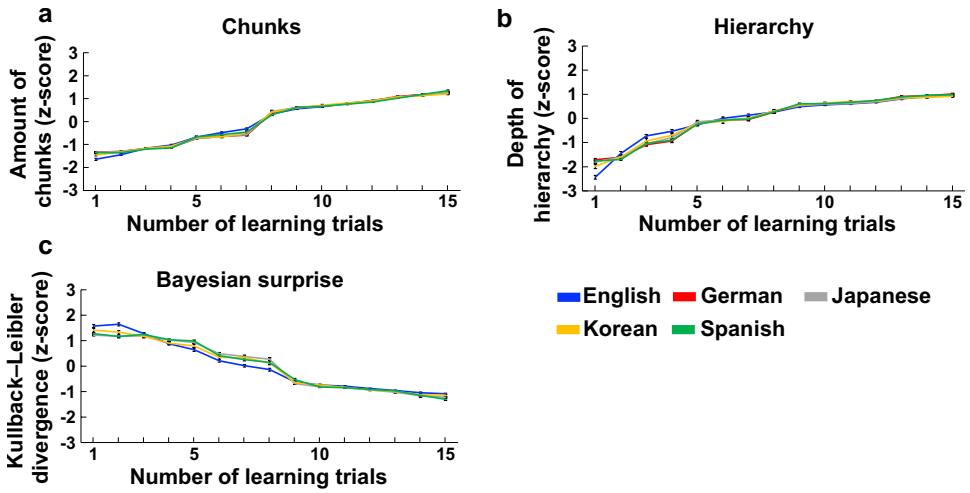

Figure 2. Statistical learning processes in each trial of learning. The numbers of chunks (**a**), the numbers (or depth) of hierarchy (**b**) and the total Bayesian surprise (or prediction error) (**c**) through each trial of statistical learning. Blue, red, grey, yellow and green represent children's songs in English, German, Japanese, Korean and Spanish, respectively. Each value was normalized to the z score. The error bars indicate the standard error of the mean.

Production process

Acoustic dimension

Using the songs composed after each statistical learning trial, the acoustic property (amplitude modulation waveforms) below 15 Hz, which corresponds to auditory rhythm, were extracted using the Bayesian model called PAD[20]. All results were shown in an external source (https://osf.io/cqmz8/?view_only=b95d94626a364700adb9e1e94384525d). The modulators were converted into time–frequency domains using a scalogram. Then, the average frequency power among the 20 songs generated by each model was calculated. The results showed that the composed songs gradually increased lower-band rhythm (1–3 Hz, 1 or 2 Hz peak) and decreased higher-band rhythm (3–5 Hz, 4 Hz peak) (Fig. 3) although the effect seems to be weak. Furthermore, over the course of 15 statistical learning trials, new music was generated gradually in comparison to the original songs (i.e. the training songs). These findings were observed consistently across the cultures studied.

Probabilistic dimension

The average probability distribution of the 20 songs generated by a model after each trial of statistical learning was calculated. The statistical similarity of the probability distributions of music composed after each of the 15 trials of statistical learning was compared to the training data (i.e. original data) using tSNE. The results showed that novel music was generated gradually in comparison to the original songs (i.e. the training songs) over the course of 15 statistical learning trials (Fig. 4). These findings were observed consistently across the cultures studied.

Discussion

Statistical learning starts at an early age for language and music learning regardless of its culture. This study examined how children's songs are acquired through statistical learning, using the computational model mimicking the statistical learning processes of the brains. I conducted a simulation experiment to visualize the temporal dynamics of perception and production processes through statistical learning among different cultures. The results suggested that, through statistical learning, the models gradually reduced Bayesian surprise that occurred by learning, and increased the number of chunks and hierarchy of knowledge. Furthermore, in production, statistical learning leads to the weak but gradual increase of delta-band (1–3 Hz) rhythm. While the observed increase was relatively subtle, it showed consistent growth with each learning repetition (i.e. 1 to 15 times). Finally, through statistical learning, new music can be generated gradually in comparison to an original song that was learned. These findings were observed consistently across the different cultures studied. These findings indicate that the statistical learning capacity of the brain contributes to the acquisition and generation of delta-band rhythm, chunking, and hierarchy, in multiple cultures.

Statistical learning to phase entrainment

The hierarchical structure of rhythms is essential for the acquisition of auditory knowledge such as music and language[18]. Particularly, the higher hierarchy corresponds to a frequency band (approximately, 3 Hz >) of prosody, intonation, and long musical notes (e.g. half notes), while the lower hierarchy corresponds to a frequency band (approximately, 3 Hz <) of syllable and musical notes (e.g. quarter notes)[19]. Such a hierarchical structure of rhythm can be visualized from the AM envelope of sound waveforms[20,21]. Evidence has shown that neural

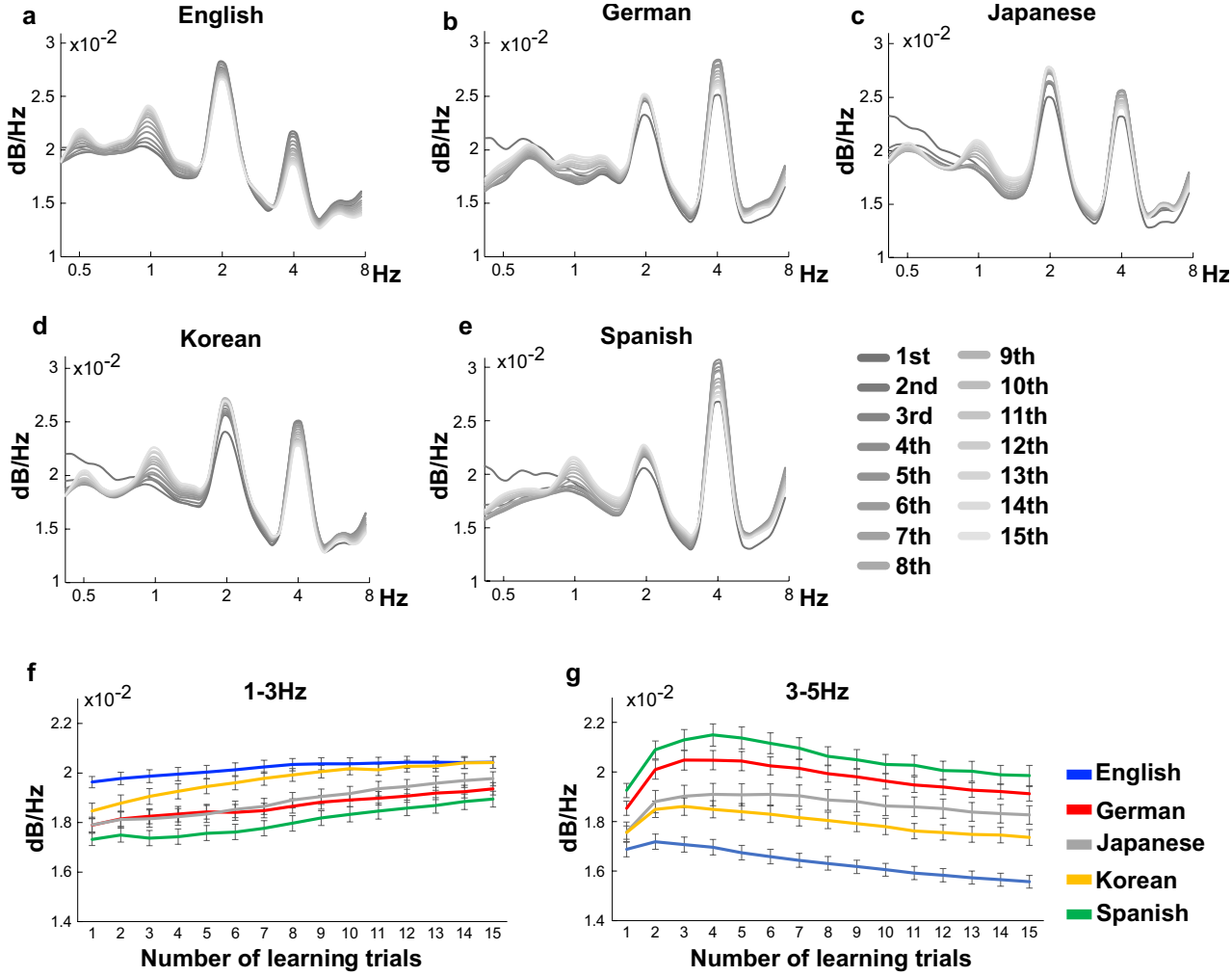

Figure 3. Acoustic properties of the composed music after each trial of statistical learning in English (**a**), German (**b**), Japanese (**c**), Korean (**d**) and Spanish (**e**) songs. The rhythm waveforms (amplitude modulation waveforms) below 15 Hz were extracted using the Bayesian model called PAD[20]). Then the modulators were converted into time–frequency domains using a scalogram. Finally, the average frequency power among the 20 songs generated by each model was calculated. The frequency powers between 1 to 3 Hz (**f**) and between 3 to 5 Hz (**g**) were averaged. Blue, red, grey, yellow and green lines represent children's songs in English, German, Japanese, Korean and Spanish, respectively.

oscillation is entrained in the AM hierarchy of auditory rhythm[23]. Such a phase entrainment contributes to the parsing of the sound signal into each unit of the rhythm hierarchy[24].

The findings in this study suggested that statistical learning may play a latent role in the acquisition of slower rhythms (< 3 Hz). A recent study by Attaheri et al.[31] has highlighted the importance of neural processing of the slower rhythm, specifically, the oscillatory phase entrainment of the delta-band rhythm, for early auditory learning. Previous studies have also demonstrated that the individual capacity for phase entrainment at the lower-frequency band is associated with their statistical learning ability[32]. Furthermore, the neural oscillations appear to synchronize with a statistically chunked rhythm acquired through statistical learning[27,28]. These findings suggest that statistical learning processes play a critical role in the acquisition and consolidation of the slower-band rhythm that forms chunked information. That is, the primary frequency of chunks formed by statistical learning (e.g. words in language or short phrases in music) is rich in the delta band frequency[27,28]. Consequently, it is suggested that through statistical learning and chunking, the brain might become entrained to the information within delta-band frequencies.

Our findings also indicated that as the slower rhythms gradually increased, the faster rhythms gradually decreased, highlighting the relationships between these rhythms. It has been suggested that the oscillators are dynamically modulated from slower to faster bands in a top-down manner[23,33]: delta oscillators modulate the theta oscillators, and theta oscillators modulate the gamma oscillators. Such a 'cascade' oscillatory system is thought to contribute to encoding the rhythm hierarchy and thus parsing of large (e.g. prosody or phrase) and smaller (e.g. syllables or tone) units in a top-down manner[23].

Auditory information directed towards children, such as speech directed towards infants and children, has been found to have a stronger rhythm in the delta range and a weaker rhythm in the theta range compared to

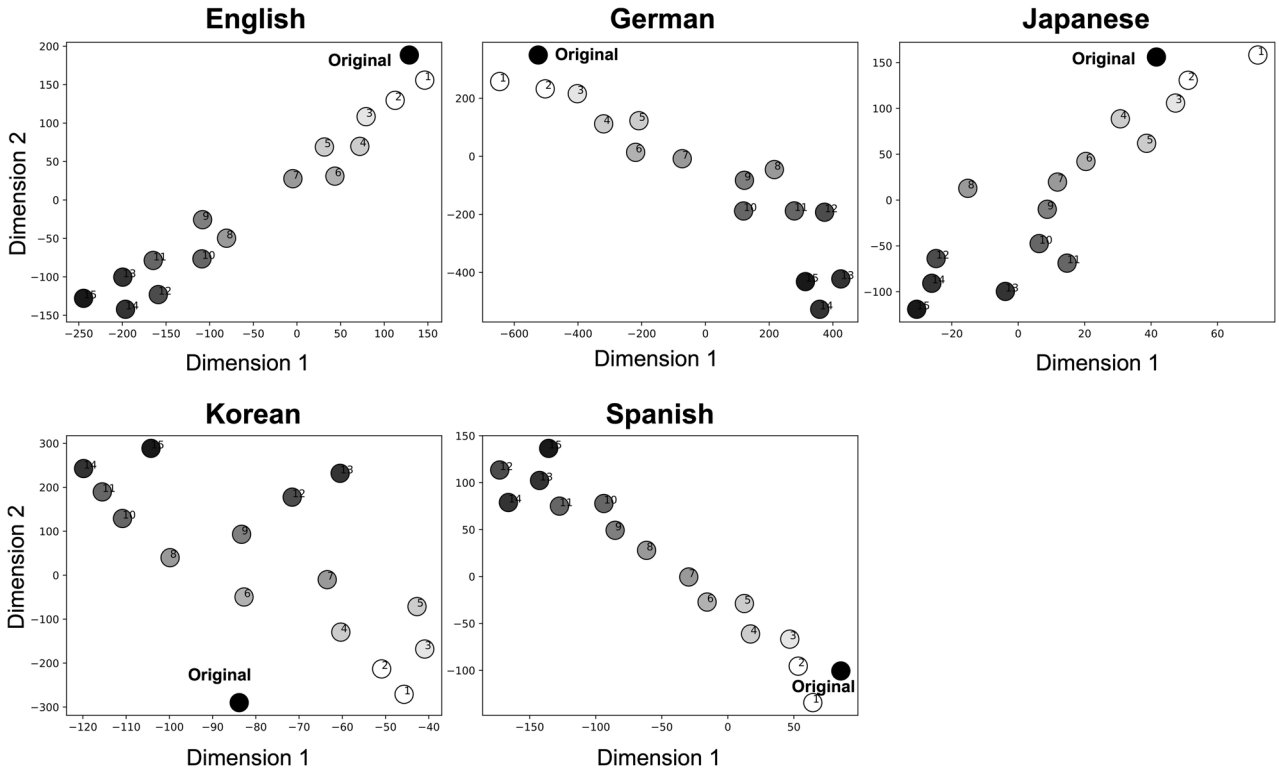

Figure 4. Statistical properties of the composed music after each trial of statistical learning. The average probability distribution of the 20 songs generated by a model after each trial of statistical learning was calculated. The statistical similarity of the probability distributions of music composed after each of the 15 trials of statistical learning was compared to the training data (i.e. original data) using t-distributed stochastic neighbour embedding (tSNE). The numbers represent the number of learning trials.

auditory information directed towards adults[34]. Further, another study comparing the hierarchical rhythmic structure across various types of sounds (nature, speech, instrumental music, song, animal sounds, etc.) has demonstrated that the slower band hierarchy (< 3 Hz) is especially pronounced in children's songs and infant/child-directed speech compared to other types of sounds[19] and across different cultures[30]. Thus, auditory stimuli directed towards children commonly exhibit a stronger power of slow rhythms. Evidence has shown that the infant's brain is first phase-entrained to the slower rhythm[31], suggesting that such a slower-band rhythm is important for early learning of language and music. Thus, it is possible that child-directed speech and songs may emphasize the slower rhythms because they syntactically and acoustically learn these rhythms first.

Prediction error and chunking

As the statistical learning progressed, it became evident that Bayesian surprise decreased across all songs and cultures. It has been theorized that the predictive processing of music works to minimize the prediction error between the bottom-up auditory signals and the top-down predictive signals based on internal music models acquired through statistical learning[2,3]. The findings in the present study may imply that statistical learning successfully worked to minimize the prediction error as well as generate chunks, which is a core function in statistical learning.

This study also detected both the increase in the number of chunks and hierarchy as the statistical learning progressed. Researchers have proposed two types of hierarchical statistical learning systems[10,16]. The first system constitutes the fundamental function of statistical learning, which groups chunks of information with high transition probabilities and integrates them into a cohesive unit. The second system involves statistical learning that arranges various chunked units to form a hierarchical syntactic structure (Fig. 1). The finding in this study may suggest that the "iteration" of statistical learning is necessary to produce many chunks, which leads to a hierarchically structured building.

The t-SNE analysis indicated that, through statistical learning, new music can be generated gradually in comparison to an original song that was learned. This suggests that statistical learning contributes to the ability to generate novel music. An individual internal model (probability distribution) that differs from the probability distribution of the learned song can be constructed by chunking through statistical learning. Such an internal model has a hierarchical structure depending on the amount of learning. This hierarchy contributes to combining different chunks that were not present in the original song. This combining may facilitate the generation of new music.

Figure 1 shows an example of such a process based on the hierarchical statistical learning of music[10]. Music has general regularities in chord sequences such as the so-called "two-five-one (II–V–I) progression." Such

syntactic progression frequently occurs in music, and therefore, the statistics of the sequential information have high transitional probability and low uncertainty. Thus, once a person has learned the statistical characteristics, it can be chunked as a commonly used unit among musicians. In contrast, the ways of combining the chunked units (e.g. combining among different blue circled units) are different between musicians. That is, the generation of chunks in statistical learning could alter the probabilistic structure of the internal model. This modification enables the creation of novel sequence information based on connections between different chunks, even when such sequences are not present in the original data. In this study, as chunk formation progresses (as shown in Fig. 2a), the created information becomes gradually distinct from the original information (Fig. 4). While novelty is not solely a component of the origin of creativity[35], it is possible that such integration of different chunked units in statistical learning may play a pivotal role in the emergence of creativity [43]. Further studies are needed to fully understand the relationship between statistical learning and creativity. The present study's findings may provide important insights to clarify this relationship.

Previous studies have shown that the brain possesses the capacity to interpret and phase-synchronize with syntactically rhythmic patterns within sound sequences, even in the absence of an acoustically rhythmic hierarchy[26]. This capability hinges on the syntactic intricacies embedded within the sequences[25]. Similarly, within the ambit of statistical learning, neural oscillations are found to synchronize with chunked rhythm, transcending the need for an overt rhythmic hierarchy in the acoustic templates of sound sequences[27,28]. This suggests the brain's ability to discern hierarchical structures of chunk rhythms, regardless of the acoustic rhythmic hierarchy[29]. Our study observed the representation of the 1–3 Hz frequency band in compositions autonomously generated subsequent to a statistical learning intervention, compared to the original song. This indicates that the brain might generate this frequency from original song as well as synchronize with it when influenced by statistical learning. Nevertheless, it's worth noting that statistical learning isn't the only factor in generating and synchronizing with the 1–3 Hz rhythm. For example, earlier research suggested the role of elements like intonation and accent in the chunking process[36].

Neurophysiological and behavioural studies have mostly examined differences in cognitive basis across music culture and experiences resulting from statistical learning[12,13]. However, it remained unknown about the temporal dynamics of the emergence of cultural systems during the learning process. Using a computational model, this study could gain a constructive understanding of such temporal dynamics including learning and production across different musical cultures.

A limitation of our study is that while we focused on children's songs from five countries, it remains unclear whether analogous results can be observed across other diverse cultures. In particular, we may not find the same outcomes in ethnomusicological traditions where principles of beat, harmony as seen in Western music are absent, or where there's no typical sheet music. Future research should aim to explore a wider range of musical cultures, including those without Western music notations.

Our findings suggested that cultural differences may not significantly modulate the statistical learning processes since statistical learning is an essential function in the human brain across different cultures. Furthermore, statistical learning of children's songs leads to the acquisition of hierarchical knowledge and the ability to generate novel music. This study provides a novel perspective on the developmental origins of creativity and the importance of statistical learning through early development.

Conclusion

The present study provided a comprehensive understanding of the role of statistical learning in the acquisition of hierarchy, chunking and slower rhythm, which is critical for acquiring language and music. It is suggested that music-cultural differences may not significantly modulate the statistical learning processes. Furthermore, statistical learning may lead to the generation of novel music. These findings have important implications for our understanding of neural processing and cognitive development, particularly in the context of auditory learning.

Materials and methods

Hierarchical Bayesian statistical learning model

This study developed a computational model, which simulates the statistical learning processes of the brain, referred to as the HSBL model[37]. The scripts of the model have been deposited to an external source (https://osf.io/cqmz8/?view_only=b95d94626a364700adb9e1e94384525d). This is a model that integrates Bayesian estimation with Markov processes using a Dirichlet distribution as a prior distribution. This model can not only calculate the transition probabilities but also determine the "reliability of the transition probabilities" from the inverse of the variance of the prior distribution of the transition probabilities. Using the normalized values of transition probabilities and reliability, this model chunks transition patterns when the product of "reliability * probability" is greater than a constant c. The constant can be decided based on the sample length and the number of learning trials. In this study, I defined $c = 5$ given the mean sample length used in this experiment (see, Table 1 supplementary). A chunked unit can be further integrated with another chunked unit, leading to the generation of a longer unit in the higher hierarchy (see Fig. 1). That is, by the cascade of chunking during statistical learning, the model gradually forms the hierarchical structure.

Materials and learning and production processes

This study generated 15 different models by manipulating the amount of learning. I used the MIDI data of 364 children's songs, which consists of 50 English, 80 German, 80 Spanish, 74 Japanese, and 80 Korean songs as the training data, and repeated the learning of the song one to fifteen times in each of 364 songs (the titles of music and the information of each song such as sample length were shown in the supplementary material). This study investigated how the probability distribution of the model is transformed over 15 trials of learning in each song.

That is, in each music culture, the model undergoes iterative learning for each song, repeating the learning process 15 times. This iterative learning is applied consistently across all songs in each culture.

Furthermore, using the probability distribution of these 15 models, 20 pieces were probabilistically generated for each model through an automatic composition process[37].

Comparison of internal representations in the model

This study compared the total Bayesian surprise (or total prediction errors) that occurred during learning, measured by the Kullback–Leibler divergence between a distribution P(x) before learning an event (e_n) and a distribution Q(x) after learning the event (e_{n+1}), as well as the total number of chunks generated during 15 trials of statistical learning. The Kullback–Leibler divergence has often been used to measure prediction error or Bayesian surprise in the framework of predictive processing of the brain[1,38,39]. It is a metric used to measure the similarity between two different probability distributions. It represents how much information is lost when one probability distribution changes into another, and since it is non-negative, a small value indicates that the two distributions are similar. Specifically, it is calculated by taking the difference between the probability density functions of the two distributions, taking the logarithm at each point, and then computing the weighted average with respect to one of the distributions. The Kullback–Leibler divergence between two probability distributions P(x) and Q(x) is calculated using the following formula:

$$D_{KL}(P||Q) = \Sigma P(i) \log (P(i)/Q(i)). \quad (1)$$

Here, P(i) and Q(i) represent the probabilities of selecting the value i according to the probability distributions P and Q, respectively. The number of chunks and hierarchy, and values of Bayesian Surprise were normalized to z score. In addition, I calculated the average probability distribution of the 20 songs generated by each model and compared the similarity of the models to the training data (i.e. original data) using t-distributed stochastic neighbour embedding (tSNE).

Comparison of acoustic properties of rhythm

We converted the MIDI data of the 20 songs generated by each model into WAV format and extracted the rhythm waveform (modulation wave) below 15 Hz using the Bayesian probabilistic amplitude demodulation model (PAD[20]). The acoustic signals were first normalised based on the z-score (mean = 0, SD = 1) in case the sound intensity influenced the spectrotemporal modulation feature. The spectrotemporal modulation of the signals was analysed using PAD to derive the dominant AM patterns. music and speech signals can be decomposed into slow-varying AM patterns and rapidly-varying carrier or frequency modulation (FM) patterns[19,40–42]. AM patterns are responsible for fluctuations in sound intensity, which are considered to be a primary acoustic feature of perceived hierarchical rhythm. On the other hand, FM patterns reflect fluctuations in spectral frequency and noise. One can separate the AM envelopes of speech signals from the FM structure using amplitude demodulation processes. The PAD model employs Bayesian inference to infer the modulators and carrier, and to identify the envelope that best fits the data and a priori assumptions. More specifically, amplitude demodulation is the process by which a signal (yt) is decomposed into a slowly varying modulator (mt) and a rapidly varying carrier (ct):

$$y_t = m_t * c_t. \quad (2)$$

PAD employs amplitude demodulation as a process of both learning and inference. Learning involves the estimation of parameters that describe distributional constraints, such as the expected timescale of variation of the modulator. Inference involves estimating the modulator and carrier from the signals based on learned or manually defined parametric distributional constraints. This information is probabilistically encoded in the likelihood function $P(y_{1:T}|c_{1:T}, m_{1:T}, \theta)$, the prior distribution over the carrier $p(c_{1:T}|\theta)$, and the prior distribution over the modulators: $p(m_{1:T}|\theta)$. Here, the notation $x_{1:T}$ represents all the samples of the signal x, ranging from 1 to a maximum value T. Each of these distributions depends on a set of parameters θ, which control factors such as the typical timescale of variation of the modulator or the frequency content of the carrier. For more detail, the parametrised joint probability of the signal, carrier, and modulator is:

$$P(y_{1:T}, c_{1:T}, m_{1:T}|\theta) = P(y_{1:T}|c_{1:T}; m_{1:T}, \theta) * p(c_{1:T}|\theta) * p(m_{1:T}|\theta). \quad (3)$$

Bayes' theorem is applied for inference, forming the posterior distribution over the modulators and carriers, given the signal:

$$P(c_{1:T}, m_{1:T}|y_{1:T}, \theta) = P(y_{1:T}, c_{1:T}, m_{1:T}|\theta) / P(y_{1:T}|\theta). \quad (4)$$

The full solution to PAD is a distribution over the possible pairs of modulators and carriers. The most probable pair of modulator and carrier given the signal is returned:

$$m*_{1:T}, c*_{1:T} = argmax\, P(c_{1:T}, m_{1:T}|y_{1:T}, \theta). \quad (5)$$

PAD utilizes Bayesian inference to estimate the most suitable modulator (i.e. envelope) and carrier that best aligns with the data and a priori assumptions. The resulting solution takes the form of a probability distribution, which describes the likelihood of a specific setting of modulator and carrier given the observed signal. Thus, PAD summarizes the posterior distribution by returning the specific envelope and carrier with the highest posterior probability, thereby providing the best fit to the data.

PAD can be run recursively using different demodulation parameters each time, producing a cascade of amplitude modulators at different oscillatory rates to form an AM. The positive slow envelope is modelled by

applying an exponential nonlinear function to a stationary Gaussian process, resulting in a positive-valued envelope with a constant mean over time. The degree of correlation between points in the envelope can be constrained by the timescale parameters of variation of the modulator (i.e. envelope), which can either be manually entered or learned from the data.

In the present study, I manually entered the PAD parameters to produce the modulators at an oscillatory band level (i.e. <10 Hz) isolated from a carrier at a higher frequency rate (>10 Hz). The carrier reflects components, including noise and pitches, for which the frequencies are much higher than those of the core modulation bands. In each sample, the modulators (envelopes) were converted into time–frequency domains using a scalogram. The scalograms depict the AM envelopes derived by recursive application of probabilistic amplitude demodulation. I then calculated the average frequency power at each frequency[43] and further averaged it over the 20 songs generated by each model.

Data availability

The scripts for the computational model (Hierarchical Bayesian Statistical Learning: HBSL), analysis, and all data including results have been deposited to an external source (https://osf.io/cqmz8/?view_only=b95d94626a364700adb9e1e94384525d). The other data was shown in supplementary data.

Received: 5 May 2023; Accepted: 20 October 2023
Published online: 23 October 2023

References

1. Friston, K. The free-energy principle: A unified brain theory?. *Nat. Rev. Neurosci.* **11**(2), 127–138. https://doi.org/10.1038/nrn2787 (2010).
2. Vuust, P., Heggli, O. A., Friston, K. J. & Kringelbach, M. L. Music in the brain. *Nat. Rev. Neurosci.* **23**(5), 287–305. https://doi.org/10.1038/s41583-022-00578-5 (2022).
3. Koelsch, S., Vuust, P. & Friston, K. Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* **23**(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006 (2019).
4. Pearce, M. T. & Wiggins, G. A. Auditory expectation: The information dynamics of music perception and cognition. *Topics Cogn. Sci.* **4**(4), 625–652 (2012).
5. Daikoku, T. Neurophysiological markers of statistical learning in music and language: Hierarchy, entropy and uncertainty. *Brain Sci.* **8**(6), 114 (2018).
6. Saffran, J. R. Statistical learning as a window into developmental disabilities. *J. Neurodev. Disord.* **10**(1), 35. https://doi.org/10.1186/s11689-018-9252-y (2018).
7. Okano, T., Daikoku, T., Ugawa, Y., Kanai, K., & Yumoto, M. Perceptual uncertainty modulates auditory statistical learning: A magnetoencephalography study. *Int. J. Psychophysiol.* **168**, 65–71 (2021).
8. Daikoku, T. Musical creativity and depth of implicit knowledge: Spectral and temporal individualities in improvisation. *Front. Comput. Neurosci.* **12**, 89. https://doi.org/10.3389/fncom.2018.00089 (2018).
9. Zioga, I., Harrison, P. M., Pearce, M. T., Bhattacharya, J. & Luft, C. D. B. From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity. *NeuroImage* **206**, 116311 (2020).
10. Daikoku, T., Wiggins, G. A. & Nagai, Y. Statistical properties of musical creativity: Roles of hierarchy and uncertainty in statistical learning. *Front. Neurosci.* **15**, 640412. https://doi.org/10.3389/fnins.2021.640412 (2021).
11. Schön, D. & François, C. Musical expertise and statistical learning of musical and linguistic structures. *Front. Psychol.* **2**, 167 (2011).
12. Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C. & Pantev, C. Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia* **50**(2), 341–349 (2012).
13. Daikoku, T. & Yumoto, M. Musical expertise facilitates statistical learning of rhythm and the perceptive uncertainty: A cross-cultural study. *Neuropsychologia* **146**, 107553. https://doi.org/10.1016/j.neuropsychologia.2020.107553 (2020).
14. Daikoku, T., Yatomi, Y. & Yumoto, M. Implicit and explicit statistical learning of tone sequences across spectral shifts. *Neuropsychologia* **63**, 194–204 (2014).
15. Daikoku, T., Yatomi, Y. & Yumoto, M. Statistical learning of music- and language-like sequences and tolerance for spectral shifts. *Neurobiol. Learn. Mem.* **118**, 8–19. https://doi.org/10.1016/j.nlm.2014.11.001 (2015).
16. Altmann, G. T. Abstraction and generalization in statistical learning: implications for the relationship between semantic types and episodic tokens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**(1711), 20160060. https://doi.org/10.1098/rstb.2016.0060 (2017).
17. Patel, A. D. Language, music, syntax and the brain. *Nat. Neurosci.* **6**(7), 674–681. https://doi.org/10.1038/nn1082 (2003).
18. Goswami, U. A neural basis for phonological awareness? An oscillatory temporal-sampling perspective. *Curr. Dir. Psychol. Sci.* **27**(1), 56–63 (2017).
19. Daikoku, T. & Goswami, U. Hierarchical amplitude modulation structures and rhythm patterns: Comparing western musical genres, song, and nature sounds to Babytalk. *PLoS One* **17**(10), e0275631. https://doi.org/10.1371/journal.pone.0275631 (2022).
20. Turner, R. E. & Sahani, M. Demodulation as probabilistic inference. *IEEE Trans. Audio Speech Lang. Process.* **19**(8), 2398–2411 (2011).
21. Leong, V. & Goswami, U. Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PloS One* **10**(12), e0144411 (2015).
22. Ding, N. *et al.* Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* **81**(B), 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011 (2017).
23. Gross, J. *et al.* Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLOS Biol.* **11**(12), e1001752 (2013).
24. Poeppel, D. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**(1), 245–255 (2003).
25. Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* **19**(1), 158–164 (2016).
26. Nozaradan, S., Peretz, I. & Mouraux, A. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* **32**(49), 17572–17581 (2012).
27. Batterink, L. J. & Paller, K. A. Online neural monitoring of statistical learning. *Cortex* **90**, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004 (2017).
28. Smalle, E. H. M., Daikoku, T., Szmalec, A., Duyck, W. & Möttönen, R. Unlocking adults' implicit statistical learning by cognitive depletion. *Proc. Natl. Acad. Sci. U.S.A.* **119**(2), e2026011119. https://doi.org/10.1073/pnas.2026011119 (2022).
29. Henin, S. *et al.* Learning hierarchical sequence representations across human cortex and hippocampus. *Sci. Adv.* **7**(8), eabc4530 (2021).

30. Pérez-Navarro, J., Lallier, M., Clark, C., Flanagan, S. & Goswami, U. Local temporal regularities in child-directed speech in Spanish. *J. Speech Lang. Hear. Res.* https://doi.org/10.1044/2022_JSLHR-22-00111 (2022).
31. Attaheri, A. *et al.* Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *Neuroimage* **247**, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698 (2022).
32. Assaneo, M. F. *et al.* Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat. Neurosci.* **22**(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z (2019).
33. Lakatos, P. *et al.* An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* **94**(3), 1904–1911 (2005).
34. Leong, V., Kalashnikova, M., Burnham, D. & Goswami, U. The temporal modulation structure of infant-directed speech. *Open Mind* **1**(2), 78–90. https://doi.org/10.1162/OPMI_a_00008 (2017).
35. Robert, J. S. Creativity. In *Cognitive Psychology* 6th edn (ed. Robert, J. S.) 479 (Cengage Learning, 2011).
36. Elmer, S., Valizadeh, S. A., Cunillera, T. & Rodriguez-Fornells, A. Statistical learning and prosodic bootstrapping differentially affect neural synchronization during speech segmentation. *Neuroimage* **235**, 118051 (2021).
37. Daikoku T, Nagai Y, DAIKIN Ltd. Computational algorithm to support human creativity (PAT.T:2022–077559). In Japanese. 大黒, ファン, 半田, 濵田, エンリケズ, 長井. 解析方法、統計学習システム及び解析プログラム. 特願2022–077559 (2022).
38. Baldi, P. & Itti, L. Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Netw.* **23**(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007 (2010).
39. Itti, L. & Baldi, P. Bayesian surprise attracts human attention. *Vis. Res.* **49**(10), 1295–306. https://doi.org/10.1016/j.visres.2008.09.007 (2009).
40. Elliott, T. M. & Theunissen, F. E. The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* **5**(3), e1000302. https://doi.org/10.1371/journal.pcbi.1000302 (2009).
41. Turner R. Statistical models for natural sounds. Ph.D. Dissertation, (University College London, 2010).
42. Daikoku, T., Kumagaya, S., Ayaya, S. & Nagai Y. Non-autistic persons modulate their speech rhythm while talking to autistic individuals. *Plos one.* **18**(9), e0285591 (2023).
43. Daikoku, T., Kamermans, K., & Minatoya, M. Exploring cognitive individuality and the underlying creativity in statistical learning and phase entrainment. *EXCLI J.* **22**, 828 (2023).

Acknowledgements

This research was supported by JSPS KAKENHI (22KK0157; 22H05210; 21H05063) and JST Moonshot Goal 9 (JPMJMS2296), Japan. The funding sources had no role in the decision to publish or prepare the manuscript.

Author contributions

T.D. conceived the method of data analysis and collected the behavioural data. T.D. analysed the data, prepared the figures, and wrote the manuscript and finalized the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-45493-6.

Correspondence and requests for materials should be addressed to T.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.