Neurobiology of Learning and Memory 118 (2015) 8–19

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme

Statistical learning of music- and language-like sequences and tolerance for spectral shifts

Tatsuya Daikoku, Yutaka Yatomi, Masato Yumoto [*]

Department of Clinical Laboratory, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

ARTICLE INFO

Article history:
Received 8 January 2014
Revised 25 September 2014
Accepted 2 November 2014
Available online 8 November 2014

Keywords:
Language and music
Statistical learning
Domain generality
Spectral shifts
Markov process
Magnetoencephalography

ABSTRACT

In our previous study (Daikoku, Yatomi, & Yumoto, 2014), we demonstrated that the N1m response could be a marker for the statistical learning process of pitch sequence, in which each tone was ordered by a Markov stochastic model. The aim of the present study was to investigate how the statistical learning of music- and language-like auditory sequences is reflected in the N1m responses based on the assumption that both language and music share domain generality. By using vowel sounds generated by a formant synthesizer, we devised music- and language-like auditory sequences in which higher-ordered transitional rules were embedded according to a Markov stochastic model by controlling fundamental ($F0$) and/or formant frequencies ($F1$–$F2$). In each sequence, $F0$ and/or $F1$–$F2$ were spectrally shifted in the last one-third of the tone sequence. Neuromagnetic responses to the tone sequences were recorded from 14 right-handed normal volunteers. In the music- and language-like sequences with pitch change, the N1m responses to the tones that appeared with higher transitional probability were significantly decreased compared with the responses to the tones that appeared with lower transitional probability within the first two-thirds of each sequence. Moreover, the amplitude difference was even retained within the last one-third of the sequence after the spectral shifts. However, in the language-like sequence without pitch change, no significant difference could be detected. The pitch change may facilitate the statistical learning in language and music. Statistically acquired knowledge may be appropriated to process altered auditory sequences with spectral shifts. The relative processing of spectral sequences may be a domain-general auditory mechanism that is innate to humans.

1. Introduction

Highly structured tone sequences, such as music and language, convey important information for human social communication. Therefore, we learn them consciously or unconsciously from birth. However, the mechanisms of learning both music and language have yet to be entirely clarified. In recent studies on music and language learning, “domain specificity” versus “domain generality” is among the issues that have received the most attention.

Domain specificity means that each type of knowledge can be innately specialized for handling tasks in a specific domain. Hauser and Chomsky et al. claimed that many aspects of language have a “universal grammar” (Hauser, Chomsky, & Fitch, 2002). In other words, these authors suggested that language has intrinsic rules that are essential for its acquisition. This viewpoint is also known as innatism. Furthermore, Jackendoff and colleagues reported that some parts of musical capacities (e.g., isochronic metrical grids, tonal pitch spaces, hierarchical tension and attraction contours based on the structure of a melody) could be acquired by domain-specific processing of music (Jackendoff & Lerdahl, 2006). Taken together, according to Hauser and Chomsky et al. and Jackendoff et al., the acquisition of language and music may require specific capacities that are independent of one another because language and music include distinct primitive structures (e.g., phonemes, syllables, phrases, syntax, and pitch classes). Therefore, learners may have to extract many independent specialized structures that are involved in each domain for the acquisition of the relevant knowledge.

In contrast, domain generality means that almost all knowledge can be innately generalized to handle tasks in all domains. Some studies have suggested that learners can extract statistical regularities that are involved in all domains (i.e., statistical learning). Therefore, statistical learning is conceived as the domain-general mechanism for the acquisition of any type of knowledge. Saffran, Johnson, Aslin, and Newport (1999) suggested that humans pro-

[*] Corresponding author at: Department of Clinical Laboratory, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Fax: +81 3 5689 0495.

E-mail address: yumoto-tky@umin.ac.jp (M. Yumoto).

http://dx.doi.org/10.1016/j.nlm.2014.11.001
1074-7427/© 2014 Elsevier Inc. All rights reserved.

cess a mechanism that computes transitional probabilities in auditory sequences, such as spoken languages and musical melodies. Furthermore, previous studies have suggested that statistical learning of higher-order structures can be interpreted by the Bayesian or the Markov models, regardless of the sensory modalities such as sight and hearing (Furl et al., 2011; Kiebel, Daunizeau, & Friston, 2008; Rao, 2005).

Conversely, Saffran et al. (1999) also stated that the acquisition of specific knowledge, such as syllables in language, might require domain-specific mechanisms that are different from statistical learning. In other words, statistical learning is not sufficient to account for all levels of the language-learning process. Previous studies have also suggested two steps of a hierarchical learning process (Ellis, 2009; Jusczyk, 1999; Saffran et al., 1999). The first step is statistical learning, which has a common mechanism among all the domains (domain generality). The second step is more highly structured specific learning, which has different mechanisms in each domain, such as linguistic syllables and musical tonality, among others (domain specificity). However, the notion of a hierarchical learning process also implies that, regardless of the domain, statistical learning plays an essential role, at least in an earlier step of the learning process (Archibald & Joanisse, 2013). In the present study, we investigated the statistical learning of three types of auditory sequences, as reflected in the neuromagnetic responses.

In recent studies of learning, it has been revealed that learners' brain activity can depend on the knowledge that they have already acquired (Furl et al., 2011; Yumoto et al., 2005). This finding suggests that if the regularities that learners have already acquired are used in an experiment, the level of learning achievement cannot be exactly equal across all the participants. Therefore, in this study, we devised original rules in which it was unlikely that any of the participants had prior experience (Furl et al., 2011; Saffran, 2003a,b; Saffran, Aslin, & Newport, 1996; Saffran et al., 1999). Thus, we equalized the level of learning achievement in all of the participants. However, the learning regularities that people have never experienced and therefore cannot predict are merely the first step of learning. For example, in most learning activities of healthy humans, newly encountered regularities are also recognized in relation to the regularities that have already been acquired. This idea has also been supported by studies that are based on information theory (Olshausen & Field, 2004). If we separately recognize all of the information that we receive as entirely different information, we must acquire a very large amount of information. This process is systematically redundant. Conversely, if we can "relatively" recognize new information by correlating it with other information that we have already learned, then we do not have to accumulate all of the received information and can spare memory capacity in our brains. This process is systematically efficient.

The first example of this efficient processing is relative pitch processing. Listeners can easily recognize transposed melodies if they have already listened to the original melodies. Even infants can possess relative pitch (Plantinga & Trainor, 2005; Trehub, 2001; Trehub & Hannon, 2006). The second example of efficient processing is the recognition of spoken languages. When we hear two sounds that are both in the same phonetic category, we can generalize the two sounds as the same phoneme due to their acoustic features such as formant frequencies (Houston & Jusczyk, 2000; Trainor, Wu, & Tsang, 2004). Houston and colleagues reported that, by 10.5 months, infants could generalize words across talkers of both sexes (Houston & Jusczyk, 2000). These processing efficiencies were available despite substantial differences in other acoustic parameters, such as pitch and harmonics. In other words, we can categorize a large amount of information by referring to common regularities in the information (e.g., relative patterns of melody or formant frequencies).

In the present study, we investigated whether the participants could relatively and efficiently learn the transitional probabilities. This approach allowed us to clarify whether the relative processing is also available in domain-general statistical learning. Saffran suggested that the relative processing of fundamental frequencies is undertaken during statistical learning (Saffran, 2003a). However, it has not yet been clarified whether the relative processing of the formant frequencies has also been performed. To further clarify the detailed mechanisms of statistical learning, it is important to determine the relative and efficient processing abilities of humans.

Mismatch negativities (MMNs) that peak at approximately 100–250 ms after the onset of deviant stimuli have been extensively studied as indices for differential processing of probable and improbable tones (i.e., standards and oddballs) (Haenschel, Vernon, Dwivedi, Gruzelier, & Baldeweg, 2005; Shestakova et al., 2002, Ross et al., 2009, Näätänen, Paavilainen, Rinne, & Alho, 2007). In previous studies on statistical learning using word segmentation paradigms, the amplitudes of the N1 responses to tones that appeared with lower probability were increased compared with responses to tones that appeared with higher transitional probability (Abla, Katahira, & Okanoya, 2008; Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012). Thus, the N1 responses have been used as an index of statistical learning of auditory sequences. Paraskevopoulos and colleagues also suggested that modulation of neural responses in the latency range of N1m (the magnetic counterpart of the N1 potential) during statistical learning could be interpreted as the MMN.

In several fields of study, such as natural language processing (Poon & Domingos, 2007; Poon & Domingos, 2008; Singla & Domingos, 2006), music perception and statistical learning (Daikoku, Yatomi, & Yumoto, 2014; Richardson & Domingos, 2006), the Markov chain has often been used as a model of the artificial grammar of language and music. The Markov chain, which was first reported by Markov (1971), is a mathematical system in which the probability of the forthcoming state is statistically defined only by the latest state. The use of the Markov chains embedded in tone sequences allows us to verify the mechanism of statistical learning in the acquisition of language and music more intrinsically than conventional oddball sequences. In the present study, using tone sequences based on second-order Markov chains that have more general structure than the word segmentation task, we verified that both the statistical learning and the relative pitch processing of auditory sequences involved in music and language are reflected in the N1m responses.

2. Methods

2.1. Participants

Fourteen right-handed (Edinburgh handedness questionnaires; laterality quotient ranged from 57.9 to 100; Oldfield, 1971) healthy Japanese participants were included. According to self-reports, they had no history of neurological or audiological disorders (eight males, six females; age range: 24–36 years). None of the participants had experience with living abroad, and none of the participants possessed absolute pitch according to self-reports. This study was approved by the Ethics Committee of The University of Tokyo. All of the participants were well informed of the purpose, safety and protection of personal data in this experiment, and they provided written informed consent for this study.

2.2. Stimuli

2.2.1. Tones

Using a cascade-Klatt type synthesizer (Klatt, 1980) HLsyn (Sensimetrics Corporation, Malden, MA, USA), we generated complex

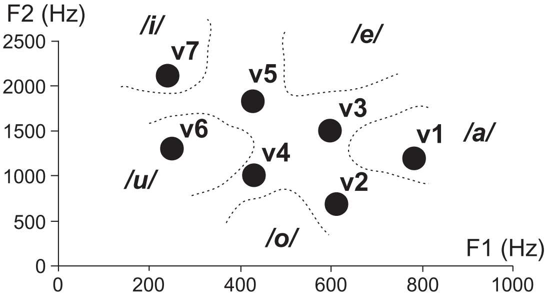


Fig. 1. The tones in the *F*1–*F*2 space used in the present study. Seven different combinations of *F*1 and *F*2 are assigned as *v*1 (780, 1200 Hz), *v*2 (610, 700 Hz), *v*3 (600, 1500 Hz), *v*4 (430, 1000 Hz), *v*5 (420, 1800 Hz), *v*6 (250, 1300 Hz), and *v*7 (240, 2100 Hz). The vowel set "*v*1, *v*2, *v*3, *v*4, *v*5" or "*v*3, *v*4, *v*5, *v*6, *v*7" sounds similar to /a/, /o/, /e/, /u/, /i/, respectively, to Japanese listeners unless the two sets are presented simultaneously. The dashed lines delineate borders at a 90% identification rate for each vowel (Ueda and Watanabe, 1987).

tones that had seven fundamental frequencies (F0) in a five-tone equal temperament ($F0 = 100 \times 2^{(n-1)/5}$ Hz, $n = 1$–7; 100, 115, 132, 152, 174, 200, and 230 Hz) and seven different combinations of the first (*F*1) and second (*F*2) formant frequencies: *v*1 (780, 1200 Hz), *v*2 (610, 700 Hz), *v*3 (600, 1500 Hz), *v*4 (430, 1000 Hz), *v*5 (420, 1800 Hz), *v*6 (250, 1300 Hz), and *v*7 (240, 2100 Hz), which sound similar to two distinct sets of five vowels to Japanese speakers when *v*1–5 or *v*3–7 are presented in isolation (Fig. 1). Only *F*0, *F*1 and *F*2 were variable, whereas all of the other parameters were constant ($F3 = 2500$ Hz, $F4 = 3500$ Hz, $F5 = 4500$ Hz, duration 400 ms, including a rise/fall of 10/200 ms, binaural presentation with the intensity of 80 dBSPL; Table 1). The *F*1 and *F*2 of every tone were chosen within a frequency range of vowels that a healthy human can generate (Ito, Tsuchida, & Yano, 2001; Kewley-Port & Watson, 1994; Peterson & Barney, 1952). The relative intensities of the formant frequencies with respect to the fundamental frequencies were defined by the default parameters of the HLsyn: the area of the glottis 4; the area at the lips 100; the area of the tongue blade 100; the area of the velopharyngeal 0; and active vocal-tract expansion 0.

2.2.2. Sequences

We devised three types of tone sequences: a music-like sequence (pitch sequence: *F*0 variable, *F*1, *F*2 constant), language-like sequence without a pitch change (formant sequence: *F*0 constant, *F*1, *F*2 variable), and language-like sequence with a pitch change (pitch-formant sequence: *F*0, *F*1, and *F*2 variable). Every sequence consisted of 750 tones separated by constant inter-stimulus intervals (ISIs) of 200 ms. Every sequence could be divided by three into the first, middle, and last portions, each with 250 tones (Fig. 2).

In the pitch sequence (Table 2, top), the first and middle portions consisted of vowel tones *v*4 (Fig. 1) with the five *F*0s ($F0 = 100 \times 2^{(n-1)/5}$ Hz, $n = 1$–5). The last portion consisted of vowel tones *v*4 with the *F*0s shifted from $n = 1$–5 to 3–7. In the formant sequence (Table 2, middle), the first and middle portions consisted of vowel tones *v*1–5 with the *F*0 of 152 Hz ($n = 4$). The last portion consisted of vowel tones shifted from *v*1–5 to *v*3–7. In the pitch-formant sequence (Table 2, bottom), the first and middle portions consisted of vowel tones *v*1–5 with the *F*0s of $n = 1$–5, respectively. The last portion consisted of vowel tones with the *F*0s and *F*1–*F*2s shifted from $n = 1$–5 to 3–7 and *v*1–5 to *v*3–7, respectively. The experimental order of the three sessions of sequences was counterbalanced across the participants. In all of the tone sequences, a one-second silent period was pseudo-randomly inserted (i.e., ISI 1.2 s) within every set of 30 successive tones. During the experiments, the participants were instructed to raise their right hands at every silent period to confirm that they were continuing to pay attention to the tone sequences.

The order of the tones in every sequence was defined by a second-order Markov process with the constraint that the probability of a forthcoming tone was statistically defined by the latest two successive tones. Fig. 3 shows the three transitional matrices of the Markov chains used in the present study. Each pair of two subsequent tones (row) could be followed by one of the fives tones

Table 1
The acoustic parameters of the presented tones in each sequence. The f_C and f_{BW} represent the center frequency and bandwidth of the formant frequencies, respectively. The (n) is the variable in $F0 = 100 \times 2^{(n-1)/5}$ Hz. The *v*1–*v*7 correspond to the tones in the *F*1–*F*2 space in Fig. 1.

	*v*4 (n = 1)	*v*4 (n = 2)	*v*4 (n = 3)	*v*4 (n = 4)	*v*4 (n = 5)	*v*4 (n = 6)	*v*4 (n = 7)
(a) Pitch sequence							
*F*0	100 (0)	115 (0)	132 (0)	152 (0)	174 (0)	200 (0)	230 (0)
*F*1	430 ± 40 (9)	430 ± 40 (7)	430 ± 40 (7)	430 ± 40 (8)	430 ± 40 (9)	430 ± 40 (14)	430 ± 40 (12)
*F*2	1000 ± 45 (−6)	1000 ± 45 (−6)	1000 ± 45 (−6)	1000 ± 45 (−4)	1000 ± 45 (−9)	1000 ± 45 (−5)	1000 ± 45 (−1)
*F*3	2500 ± 75 (−32)	2500 ± 75 (−32)	2500 ± 75 (−32)	2500 ± 75 (−31)	2500 ± 75 (−30)	2500 ± 75 (−29)	2500 ± 75 (−32)
*F*4	3500 ± 175 (−46)	3500 ± 175 (−46)	3500 ± 175 (−46)	3500 ± 175 (−45)	3500 ± 175 (−44)	3500 ± 175 (−44)	3500 ± 175 (−44)
*F*5	4500 ± 250 (−56)	4500 ± 250 (−58)	4500 ± 250 (−57)	4500 ± 250 (−56)	4500 ± 250 (−55)	4500 ± 250 (−55)	4500 ± 250 (−55)
					$f_C \pm 1/2\ f_{BW}$ (relative intensity (dB))		
	*v*1 (n = 4)	*v*2 (n = 4)	*v*3 (n = 4)	*v*4 (n = 4)	*v*5 (n = 4)	*v*6 (n = 4)	*v*7 (n = 4)
(b) Formant sequence							
*F*0	152 (0)	152 (0)	152 (0)	152 (0)	152 (0)	152 (0)	152 (0)
*F*1	780 ± 40 (15)	610 ± 40 (11)	600 ± 40 (13)	430 ± 40 (8)	420 ± 40 (5)	250 ± 40 (2)	240 ± 40 (1)
*F*2	1200 ± 45 (9)	700 ± 45 (2)	1500 ± 45 (0)	1000 ± 45 (−4)	1800 ± 45 (−12)	1300 ± 45 (−20)	2100 ± 45 (−20)
*F*3	2500 ± 75 (−18)	2500 ± 75 (−38)	2500 ± 75 (−17)	2500 ± 75 (−31)	2500 ± 75 (−16)	2500 ± 75 (−36)	2500 ± 75 (−24)
*F*4	3500 ± 175 (−31)	3500 ± 175 (−49)	3500 ± 175 (−31)	3500 ± 175 (−45)	3500 ± 175 (−32)	3500 ± 175 (−50)	3500 ± 175 (−43)
*F*5	4500 ± 250 (−42)	4500 ± 250 (−61)	4500 ± 250 (−42)	4500 ± 250 (−56)	4500 ± 250 (−45)	4500 ± 250 (−62)	4500 ± 250 (−56)
					$f_C \pm 1/2\ f_{BW}$ (relative intensity (dB))		
	*v*1 (n = 1)	*v*2 (n = 2)	*v*3 (n = 3)	*v*4 (n = 4)	*v*5 (n = 5)	*v*6 (n = 6)	*v*7 (n = 7)
(c) Pitch-formant sequence							
*F*0	100 (0)	115 (0)	132 (0)	152 (0)	174 (0)	200 (0)	230 (0)
*F*1	780 ± 40 (9)	610 ± 40 (8)	600 ± 40 (6)	430 ± 40 (8)	420 ± 40 (9)	250 ± 40 (0)	240 ± 40 (0)
*F*2	1200 ± 45 (7)	700 ± 45 (9)	1500 ± 45 (−4)	1000 ± 45 (−4)	1800 ± 45 (−5)	1300 ± 45 (−20)	2100 ± 45 (−23)
*F*3	2500 ± 75 (−18)	2500 ± 75 (−37)	2500 ± 75 (−16)	2500 ± 75 (−31)	2500 ± 75 (−14)	2500 ± 75 (−36)	2500 ± 75 (−30)
*F*4	3500 ± 175 (−34)	3500 ± 175 (−52)	3500 ± 175 (−32)	3500 ± 175 (−45)	3500 ± 175 (−32)	3500 ± 175 (−52)	3500 ± 175 (−48)
*F*5	4500 ± 250 (−45)	4500 ± 250 (−62)	4500 ± 250 (−44)	4500 ± 250 (−56)	4500 ± 250 (−44)	4500 ± 250 (−63)	4500 ± 250 (−60)
					$f_C \pm 1/2\ f_{BW}$ (relative intensity (dB))		

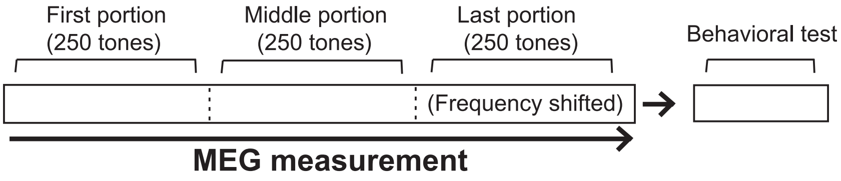


Fig. 2. Experimental design.

Table 2
Frequency parameters of tones used in the first, middle, and last portions.

		First, middle portion (500 tones)	Last portion (250 tones)
Pitch sequence (*F*0: variable)	*F*0	$100 \times 2^{(n-1)/5}$ Hz, $n = 1–5$	Frequency shifted to $n = 3–7$
	*F*1	430, 1000 Hz (*v*4[a])	
	*F*2		
Formant sequence (*F*1, *F*2: variable)	*F*0	152 Hz ($100 \times 2^{(n-1)/5}$, $n = 4$)	
	*F*1	Five vowels: *v*1–5[a]	Frequency shifted to *v*3–7[a]
	*F*2		
Pitch-formant sequence (*F*0, *F*1, *F*2: variable)	*F*0	$100 \times 2^{(n-1)/5}$ Hz, $n = 1–5$	Frequency shifted to $n = 3–7$
	*F*1	Five vowels: *v*1–5[a]	Frequency shifted to *v*3–7[a]
	*F*2		

[a] Refer to Fig. 1.

(columns). The probability for one of the five tones was 80% and 5% for the remaining four tones. Henceforth, the tones that appeared with higher and lower transitional probabilities are termed frequent and rare tones, respectively. The three different Markov chains shown in Fig. 3 were adopted in each sequence in a way that specific transitional patterns did not interfere with learning in the adjacent experimental sessions, and the order of the adoption was counterbalanced across the participants.

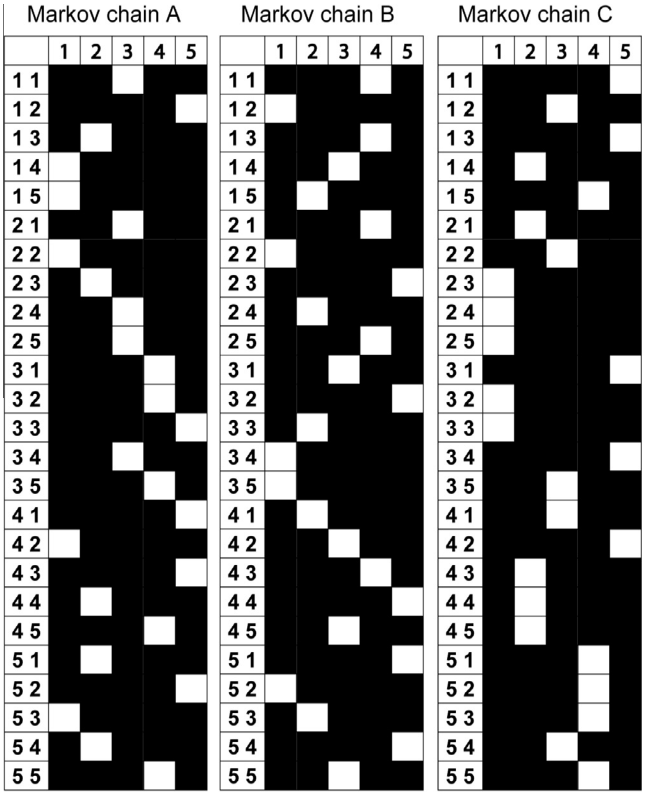
We recorded magnetoencephalographic (MEG) signals from the participants while they listened to the tone sequences.

2.3. Behavioral test

After each sequence of the MEG measurement, the participants were presented ten tone series with eight tones, half of which were sequenced by the same Markov process. The participants were asked in interviews whether each tone series sounded familiar to them. If a participant had adequately learned the transitional probabilities of the sequences, he or she should have been able to correctly answer this question. This behavioral test was completed within two minutes for each participant. After the behavioral tests, we performed the Shapiro–Wilk test for normality of the behavioral data. Depending on the result of the test for normality, either one-way repeated measures analysis of variance (ANOVA) or Friedman ANOVA was applied to the ratio of correct answers in each session (pitch, formant, and pitch-formant sessions). If significant effects were detected, the Bonferroni-corrected post hoc tests were conducted for further analysis. The statistical significance was set at $p = 0.05$ in all of the analyses.

2.4. Measurements

The auditory stimuli were sequenced with the STIM2 system (Compumedics Neuroscan, El Paso, TX, USA) and were delivered binaurally to each participant's ears at 80 dBSPL through ER-3A earphones (Etymotic Research, Elk Grove Village, IL, USA). We recorded MEG signals from the participants while they listened to the tone sequences in a magnetically shielded room. We used a 306-channel neuromagnetometer system (Elekta Neuromag Oy, Helsinki, Finland), which has 204 planar first-order gradiometers and 102 magnetometers at 102 measuring sites on a helmet-shaped surface that covers the entire scalp. Auditory stimulus-triggered epochs were filtered online with a 0.1–200 Hz band-pass filter and were then recorded at a sampling rate of 600 Hz. Using the time-domain extension of signal space separation with a buffer length of 10 s and a correlation limit of 0.980, the contamination


Markov chain A

	1	2	3	4	5
1 1	■	■	□	■	■
1 2	■	■	■	■	□
1 3	■	□	■	■	■
1 4	□	■	■	■	■
1 5	□	■	■	■	■
2 1	■	■	□	■	■
2 2	□	■	■	■	■
2 3	■	□	■	■	■
2 4	■	■	□	■	■
2 5	■	■	□	■	■
3 1	■	■	■	□	■
3 2	■	■	■	□	■
3 3	■	■	■	■	□
3 4	■	■	□	■	■
3 5	■	■	■	□	■
4 1	■	■	■	■	□
4 2	□	■	■	■	■
4 3	■	■	■	■	□
4 4	■	□	■	■	■
4 5	■	■	■	□	■
5 1	■	□	■	■	■
5 2	■	■	■	■	□
5 3	□	■	■	■	■
5 4	■	□	■	■	■
5 5	■	■	■	□	■

Markov chain B

	1	2	3	4	5
1 1	■	■	■	□	■
1 2	□	■	■	■	■
1 3	■	■	■	□	■
1 4	■	■	□	■	■
1 5	■	□	■	■	■
2 1	■	■	■	□	■
2 2	□	■	■	■	■
2 3	■	■	■	■	□
2 4	■	□	■	■	■
2 5	■	■	■	□	■
3 1	■	■	□	■	■
3 2	■	■	■	■	□
3 3	■	□	■	■	■
3 4	□	■	■	■	■
3 5	□	■	■	■	■
4 1	■	□	■	■	■
4 2	■	■	□	■	■
4 3	■	■	■	□	■
4 4	■	■	■	■	□
4 5	■	■	□	■	■
5 1	■	■	■	■	□
5 2	□	■	■	■	■
5 3	■	□	■	■	■
5 4	■	■	■	■	□
5 5	■	■	□	■	■

Markov chain C

	1	2	3	4	5
1 1	■	■	■	■	□
1 2	■	■	□	■	■
1 3	■	■	■	■	□
1 4	■	□	■	■	■
1 5	■	■	■	□	■
2 1	■	□	■	■	■
2 2	■	■	□	■	■
2 3	□	■	■	■	■
2 4	□	■	■	■	■
2 5	□	■	■	■	■
3 1	■	■	■	■	□
3 2	□	■	■	■	■
3 3	□	■	■	■	■
3 4	■	■	■	■	□
3 5	■	■	□	■	■
4 1	■	■	□	■	■
4 2	■	■	■	■	□
4 3	■	□	■	■	■
4 4	■	□	■	■	■
4 5	■	□	■	■	■
5 1	■	■	■	□	■
5 2	■	■	■	□	■
5 3	■	■	■	□	■
5 4	■	■	□	■	■
5 5	■	■	■	□	■

Fig. 3. The transitional matrices of the Markov chains used in the present study. Rows: the two tones that were most recently presented. Columns: the next tone that might appear. Each pair of two tones (rows) could be followed by one of the five subsequent tones (columns), with 80% (white cells) or 5% (black cells) probability. In the last portion, the tone numbers in the transitional matrices were shifted from 1–5 to 3–7, but the transitional patterns were not changed.

from environmental magnetic noise was reduced (Taulu & Hari, 2009). The waveforms were filtered offline with a 2–40 Hz band-pass and averaged in the analysis window from −50 to 500 ms. In each sequence, the responses to the tones with artifacts less than 3 pT/cm for gradiometers and less than 3 pT for magnetometers were averaged to calculate single equivalent current dipoles (ECDs) for the N1m and P2m. The baseline for the waveforms in each MEG channel was defined by the mean amplitude between −100 and 0 ms.

2.5. Data analysis

The target of analysis was the responses to the three tones that commonly appeared in any portion of each sequence (i.e., $n = 3–5$ for the pitch and pitch-formant sequences and $v3–5$ for the formant and pitch-formant sequences). Otherwise, we cannot exclude the possibility that the difference detected in the recorded responses was attributed to the physical difference in the presented tone ensembles. The MEG responses to the three tones were selectively averaged according to the transitional probabilities, from the beginning of each portion until the number averaged reached 30 times (i.e., 10 times for each tone). The sources of the N1m and P2m responses were modeled separately as an ECD in each hemisphere. The ECDs for the N1m and P2m were separately calculated from the averaged responses to all 750 tones in each sequence at the peak latencies, with a goodness of fit above 80% using the same 66 temporal channels (44 gradiometers and 22 magnetometers) for each participant. The selected channel areas are shown in Fig. 4. In the N1m and P2m, the source-strength waveforms for the averaged responses in each hemisphere were calculated using the ECDs (Numminen, Salmelin, & Hari, 1999). Then, in each session (pitch, formant, and pitch-formant sessions), we performed a 2 (hemisphere: right and left) × 3 (portion: first, middle and last) × 2 (probability: high and low) ANOVA with the peak amplitude and peak latency of the source-strength waveforms. If significant effects were detected, the Bonferroni-corrected post hoc tests were conducted for further analysis. Statistical significance was set at $p = 0.05$ in all of the analyses.

3. Results

3.1. Behavioral data

The results of the Shapiro–Wilk test for normality indicated that the behavioral data in the pitch-formant session did not follow a normal distribution ($p = .019$). The results of the Mann–Whitney U test indicated that the ratio of correct answers for each type of tone series was significantly above the chance level of 50% (pitch series: $z[2] = 3.19$, $p = 0.0014$, formant series: $z[2] = 2.51$, $p = 0.037$, pitch-formant series: $t[13] = 3.21$, $p = 0.0013$). Friedman ANOVA on the ratio of correct answers in each session (pitch, formant, and pitch-formant sessions) detected a significant difference ($x^2(2) = 12.62$, $p = .0018$). The Bonferroni-corrected post hoc test revealed that the ratios of correct answers in the behavioral tests were significantly higher in the pitch and pitch-formant sessions than in the formant session (pitch session: $p = .012$, pitch-formant session: $p = .0050$) (Fig. 5). Despite the statistical significance of the behavioral tests, none of the participants could verbalize in detail the statistical knowledge that had supposedly been acquired during each measurement session.

3.2. Magnetoencephalographic data

We confirmed that all of the participants correctly raised their right hands at every silent period in the tone sequences and that they were continuing to pay attention to the tone sequences. The peak amplitudes and latencies of the N1m and P2m responses are shown in Table 3.

3.2.1. Pitch sequence

3.2.1.1. Amplitude. The main portion effect and main probability effect on the N1m peak amplitudes were significant (portion: $F[2, 26] = 3.81$, $p = .036$, probability: $F[1, 13] = 6.28$, $p = .026$). The hemisphere-portion interaction and the portion-probability interaction of the N1m peak amplitudes were significant (hemisphere-portion: $F[2, 26] = 3.80$, $p = .036$, portion-probability: $F[2, 26] = 5.88$,

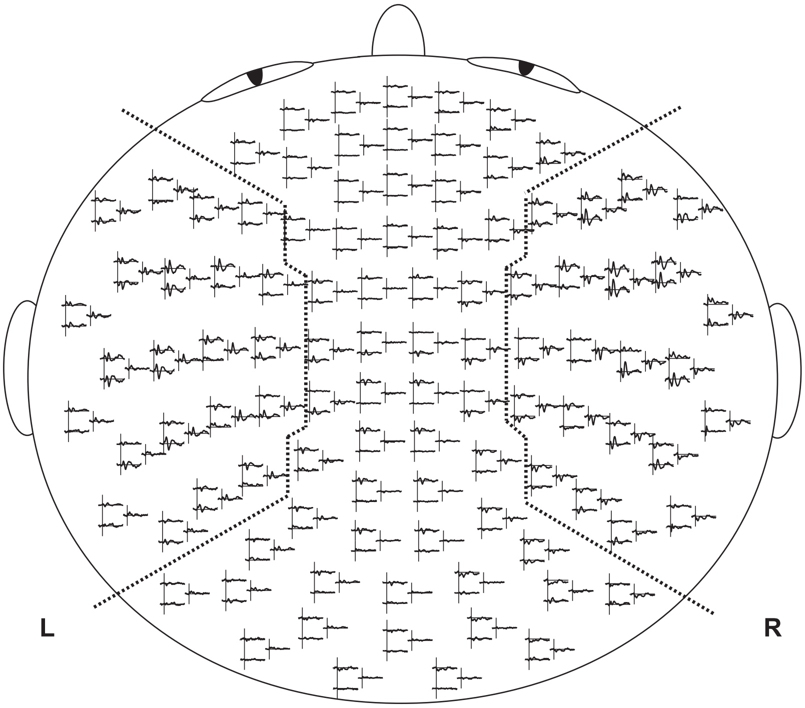


Fig. 4. The selected 66 channels in each hemisphere for the ECD and source-strength calculations are bordered with the dotted lines.

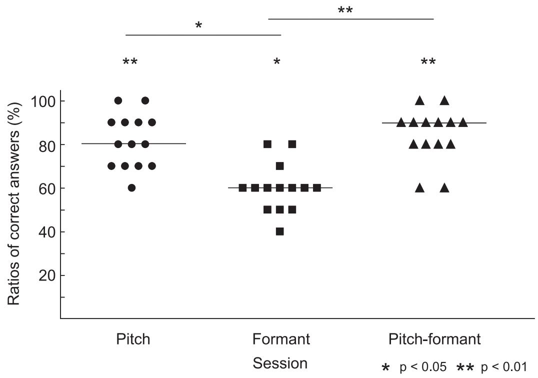


Fig. 5. The ratios of correct answers in each session ($N = 14$). The horizontal lines piercing the clusters indicate the median in each session.

$p = .008$). Moreover, further statistical analysis of the main portion effect revealed that the N1m peak amplitudes in the middle portions significantly decreased compared with those in the first portion ($p = .018$). Further statistical analysis of the hemisphere-portion interaction revealed that in the left hemisphere, the N1m peak amplitudes in the middle and last portions significantly decreased compared with those in the first portion (middle portion: $p = .029$, last portion: $p = .022$).

Further statistical analysis of the portion-probability interaction revealed the following two significant results. First, the N1m peak amplitudes for frequent tones in both the middle and last portions were significantly decreased compared with those in the first portion (middle portion: $p = .002$, last portion: $p = .009$), but no significant result was obtained for the N1m peak amplitudes for rare tones (middle portion: $p = 1.000$, last portion: $p = 1.000$). Second, in both the middle and last portions, the N1m peak amplitudes for the frequent tones were significantly decreased compared with those for the rare tones (middle portion: $p = .004$, last portion: $p = .008$) (Fig. 6).

3.2.1.2. Latency. The main portion effect and main probability effect on the N1m peak latencies were significant (portion: $F[2, 26] = 9.70$, $p = .001$, probability: $F[1, 13] = 7.27$, $p = .018$). The N1m peak latencies for the frequent tones were shorter than those for the rare tones (Fig. 7). Moreover, further statistical analysis indicated that the N1m peak latencies in the last portion were significantly longer than those in the first portion ($p = .0001$). No other significant differences in the N1m and P2m were detected in the pitch sessions.

3.2.2. Formant sequence

ANOVA revealed no significant effects in the peak amplitude and peak latency of the N1m and P2m.

3.2.3. Pitch-formant sequence

3.2.3.1. Amplitude. The main probability effect on the N1m peak amplitudes was significant ($F[1, 13] = 22.31$, $p = .0004$). The hemisphere-probability interaction of the N1m peak amplitudes were significant ($F[1, 13] = 5.64$, $p = .034$). Moreover, further statistical analysis revealed that, in both the middle and last portions, the N1m peak amplitudes for the frequent tones were significantly decreased compared with those for the rare tones (middle portion: $p = .002$, last portion: $p = .003$).

Further statistical analysis of the hemisphere-probability interaction revealed that in both the left and right hemispheres, the N1m peak amplitudes for the frequent tones were significantly decreased compared with those for the rare tones (left hemisphere: $p = .0003$, right hemisphere: $p = .006$) (Fig. 6).

3.2.3.2. Latency. The main probability effect on the N1m peak latencies was significant ($F[1, 13] = 15.20$, $p = .002$). The N1m peak

Table 3
The peak amplitudes and latencies of the N1m and P2m responses ($N = 14$).

			First portion		Middle portion		Last portion	
			High	Low	High	Low	High	Low
(a) Left hemisphere								
Pitch sequence	N1m	Amplitude	23.9 ± 3.1	21.2 ± 2.2	15.0 ± 3.7	21.0 ± 2.9	15.6 ± 2.7	20.8 ± 2.3
		Latency	111.9 ± 4.0	121.9 ± 4.5	116.0 ± 3.4	120.8 ± 3.9	124.4 ± 4.6	123.8 ± 3.9
	P2m	Amplitude	11.8 ± 2.5	13.8 ± 4.0	12.7 ± 2.2	11.4 ± 3.1	11.1 ± 2.6	10.9 ± 2.7
		Latency	242.2 ± 52.2	195.1 ± 4.5	187.0 ± 6.8	191.7 ± 6.4	192.4 ± 5.1	200.1 ± 7.6
Formant sequence	N1m	Amplitude	13.3 ± 1.5	11.6 ± 1.7	11.5 ± 1.4	13.3 ± 2.6	13.3 ± 2.5	11.3 ± 1.6
		Latency	112.9 ± 2.6	114.1 ± 2.4	117.4 ± 4.7	116.8 ± 3.6	118.2 ± 4.5	120.4 ± 4.5
	P2m	Amplitude	11.9 ± 3.6	11.2 ± 2.5	8.7 ± 2.6	9.0 ± 2.3	9.5 ± 2.5	8.7 ± 2.2
		Latency	192.3 ± 10.2	187.1 ± 6.7	199.8 ± 4.9	193.6 ± 6.9	182.1 ± 7.4	199.8 ± 9.3
Pitch-formant sequence	N1m	Amplitude	19.2 ± 2.3	23.1 ± 3.2	18.5 ± 3.1	26.3 ± 3.9	15.2 ± 2.8	22.0 ± 2.4
		Latency	111.4 ± 3.8	117.6 ± 3.7	111.6 ± 2.8	118.9 ± 2.6	109.8 ± 3.6	115.1 ± 2.5
	P2m	Amplitude	7.9 ± 2.1	9.7 ± 3.4	5.9 ± 4.5	8.2 ± 3.7	13.4 ± 6.2	8.6 ± 2.9
		Latency	196.6 ± 6.7	197.2 ± 6.0	190.2 ± 8.8	189.0 ± 6.0	181.9 ± 5.6	190.5 ± 6.9
(b) Right hemisphere								
Pitch sequence	N1m	Amplitude	21.0 ± 4.1	19.4 ± 3.5	17.7 ± 2.8	20.9 ± 3.5	19.2 ± 3.7	22.6 ± 4.0
		Latency	112.3 ± 2.8	115.6 ± 2.8	115.7 ± 2.9	120.6 ± 3.4	119.6 ± 3.8	119.9 ± 2.8
	P2m	Amplitude	17.7 ± 4.2	19.4 ± 4.6	16.5 ± 4.4	16.5 ± 3.2	17.8 ± 4.6	14.5 ± 4.1
		Latency	183.1 ± 5.9	202.7 ± 8.5	202.7 ± 9.3	193.4 ± 6.5	198.0 ± 7.2	190.2 ± 10.2
Formant sequence	N1m	Amplitude	18.3 ± 2.2	16.4 ± 2.4	14.7 ± 2.0	16.2 ± 2.5	15.0 ± 2.6	14.1 ± 1.9
		Latency	111.9 ± 2.8	115.9 ± 2.5	114.5 ± 3.8	119.9 ± 3.5	117.5 ± 4.4	119.1 ± 4.3
	P2m	Amplitude	10.9 ± 5.3	13.1 ± 3.5	12.3 ± 3.7	14.6 ± 5.8	15.9 ± 5.8	13.7 ± 5.5
		Latency	193.2 ± 6.9	193.7 ± 5.2	195.5 ± 10.7	198.6 ± 10.3	196.0 ± 10.6	196.3 ± 7.8
Pitch-formant sequence	N1m	Amplitude	23.0 ± 2.3	25.2 ± 2.5	23.4 ± 2.5	27.5 ± 2.7	21.2 ± 2.4	25.1 ± 2.6
		Latency	110.4 ± 2.9	116.9 ± 2.3	107.4 ± 2.4	117.7 ± 2.7	117.8 ± 6.9	118.9 ± 3.2
	P2m	Amplitude	14.2 ± 3.7	11.4 ± 3.1	12.1 ± 2.7	10.9 ± 3.6	16.3 ± 5.0	11.5 ± 3.1
		Latency	193.6 ± 7.0	194.0 ± 5.2	182.7 ± 6.6	197.2 ± 5.4	188.2 ± 5.7	198.6 ± 8.4

Mean ± *S.E.M* (amplitude: nAm, latency: ms).

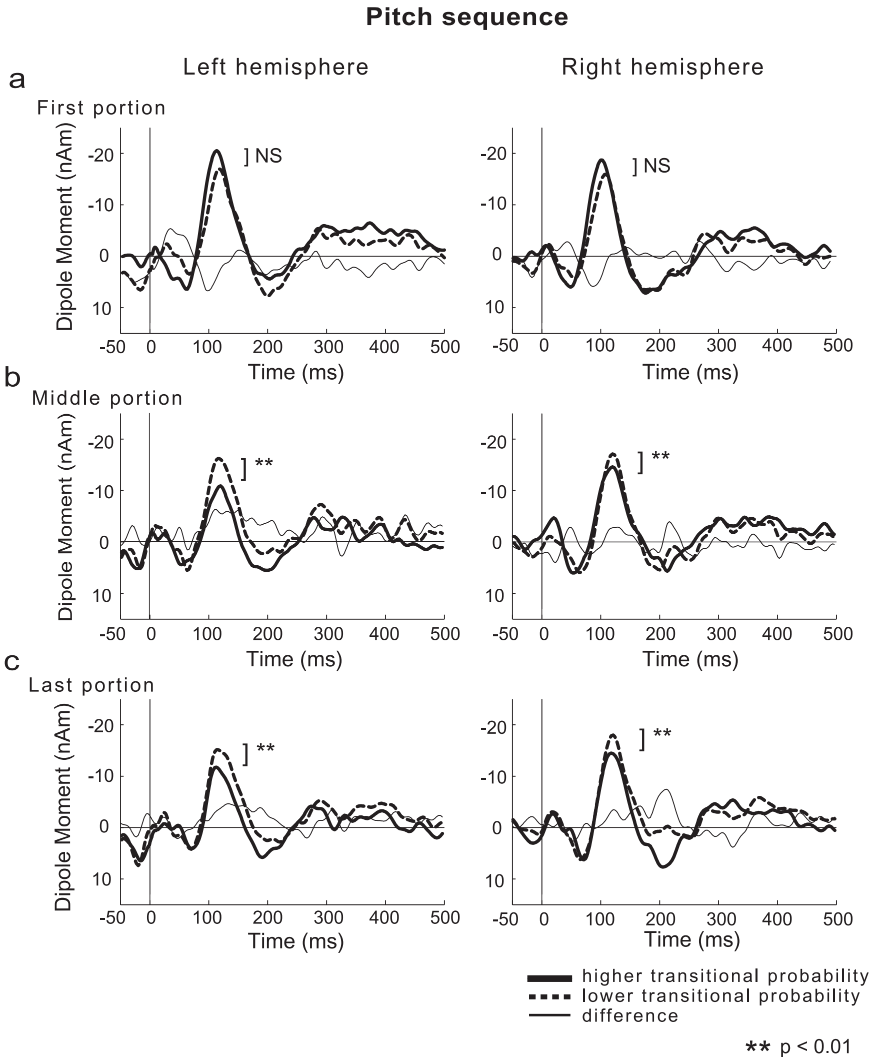


Fig. 6. Grand-averaged source-strength waveforms calculated using the ECDs modeled on the N1m responses ($N = 14$). The first (a), middle (b), and last (c) portions of the pitch sequence. The first (d), middle (e), and last (f) portions of the formant sequence. The first (g), middle (h), and last (i) portions of the pitch-formant sequence. The responses in the left and right hemispheres are located in the left and right sides, respectively. The thick lines and dashed lines indicate responses to the frequent and rare tones, respectively. The thin lines indicate the difference in waveforms between the responses to the rare and frequent tones.

latencies for the frequent tones were shorter than those for the rare tones (Fig. 7). No other significant differences in the N1m and P2m were detected in the pitch-formant sessions.

4. Discussion

If learners acquire statistical knowledge based on the transitional probabilities of the tone sequences, they can predict a tone that will follow after the preceding tones in the tone sequences. With this prediction for the upcoming tones, frequent tones (i.e., more predictable tones) lead to decreasing neural responses compared with rare tones (i.e., less predictable tones). In contrast, the violation of this prediction based on the learned transition probabilities leads to increasing neural responses compared with more predictable tones. As a result, the differences in amplitude of a learner's neural responses between the tones that appeared with the higher and lower probabilities could occur if they have learned the transitional rules of the tone sequences (Abla et al., 2008; Daikoku et al., 2014; Furl et al., 2011; Paraskevopoulos et al., 2012). Furthermore, Abla et al. also suggested that in predicting upcoming tones, frequent tones result in a shortening of the neural response latencies compared with tones that appeared with a lower probability. Ultimately, the differences in the response latency between the frequent and rare tones could occur during statistical learning.

On the other hand, neural responses to tones are attenuated with repeated stimulation due to the adaptation of auditory cortical neurons. Previous studies have reported that repeated auditory stimulation leads to the attenuation of the N1m and P2m responses (Kuriki, Ohta, & Koyama, 2007; Teismann et al., 2004). Furthermore, in the present study, the spectral frequencies of $F0$ and/or $F1$–$F2$ were relatively shifted in the last portion of all of the sequences. If the participants possess the ability to process the pitch sequences relatively, based on the transitional probabilities, the statistical learning effects should be retained even after the spectral frequencies are relatively shifted. In summary, both the learning and the adaptation effects on the amplitudes and latencies and the processing for the spectral shifts could be reflected in the neuromagnetic responses recorded in the present study. As described previously, the adults have already possessed relative pitch since childhood and acquired the ability to categorize different sounds within the same phonetic category as the same phoneme in their mother language. Therefore, the newly learned knowledge of the participants during this experiment was the sta-

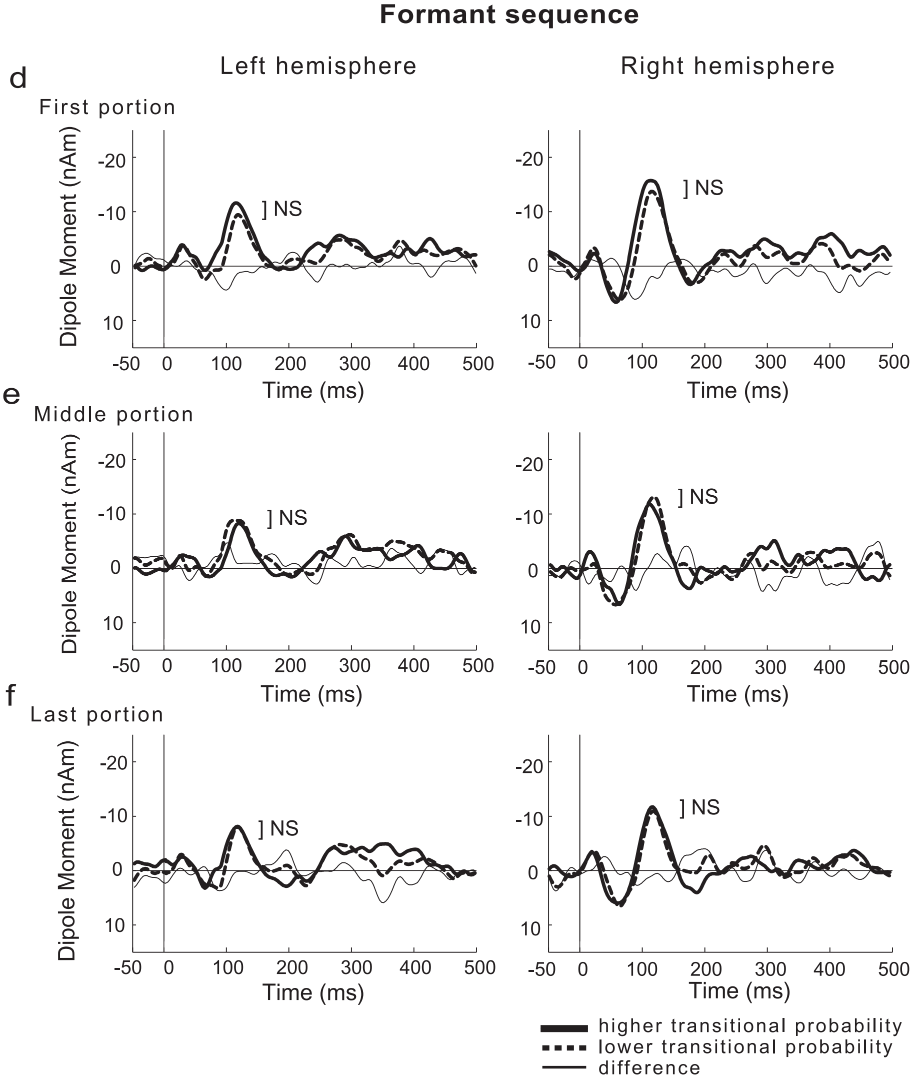


Fig. 6 (*continued*)

tistical structure that was embedded in the presented tone sequences.

In the pitch sequence, the amplitudes of the responses to the rare tones were significantly higher than those to the frequent tones, and the responses to the frequent tones were significantly attenuated in the second and last portions. Furthermore, a shortening of the neural response latencies was detected due to the potential effect of the adaptation, although the differences in the latency of the neural responses between the frequent and rare tones were not significant. Additionally, these effects of statistical learning were retained even after the spectral frequencies were relatively shifted. However, in Markov chain A (Fig. 3), frequent sequences 11–3 and 33–5 in the first and middle portions were shifted to frequent sequences 33–5 and 55–7 in the last portion, respectively. Thus, sequences 33–5 appeared frequently in all of the portions. Therefore, one cannot exclude the possibility that those sequences might have affected the processing of the spectral shifts. There were only six such cases in all of the transitional patterns (25 patterns × 3 Markov chains): four patterns (33–5, 35–4, 43–5, 45–4) in Markov chain A, no patterns in Markov chain B, and two patterns (34–5, 54–3) in Markov chain C. The three different Markov chains were adopted in each sequence, and the order of the adoption was counterbalanced across the participants. In the end, only 8% [6/(25 × 3)] of the averaged neural responses might affect the results in the last portion.

In the left hemisphere, the N1m peak amplitudes in the middle and last portions significantly decreased compared with those in the first portion. This result suggests right-hemisphere persistence in auditory responsiveness to the pitch sequence. Generally, right-handed people have left hemispheric dominance for language and right hemispheric dominance for music (Knecht et al., 2000; Schönwiesner, Rübsamen, & von Cramon, 2005). In the present study, the pitch sequence (i.e., only the pitch that could be changed) could have been processed as a musical melody, whereas the formant and pitch-formant sequences (i.e., the formant could be changed) could have been processed as spoken language. However, this result might also be interpreted as a stronger adaptation in the left hemisphere. The relationship between adaptation and persistence should be clarified in future studies.

In the pitch-formant sequence, the amplitudes of the responses to the rare tones were significantly higher than those to the frequent tones, and a shortening of the neural response latencies was detected due to the potential effect of the adaptation. Furthermore, these effects of statistical learning were retained even after the spectral frequencies were relatively shifted. Different languages with different numbers of vowels share almost the same

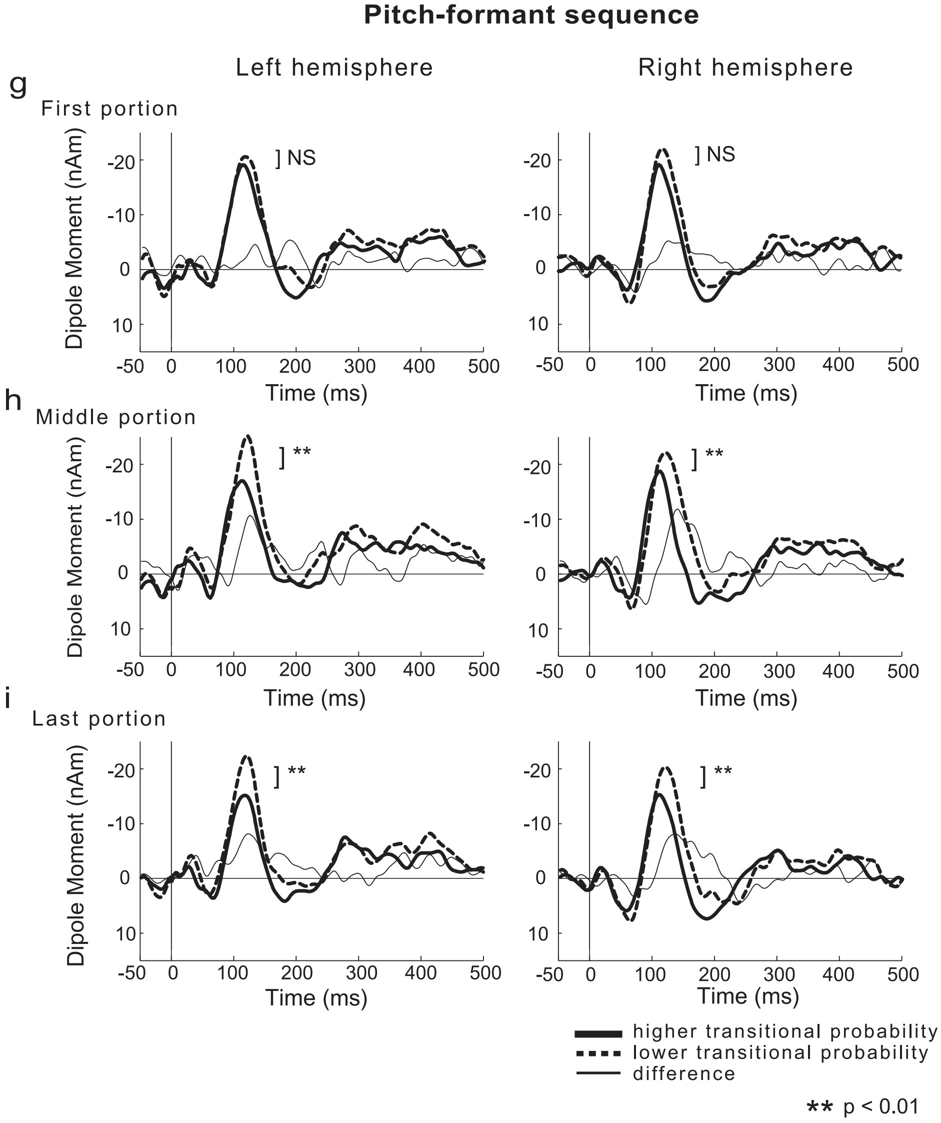


Fig. 6 (*continued*)

*F*1–*F*2 space, considering the common vocal apparatus across different races. The Japanese language has only five vowels. Therefore, the receptive area for each Japanese vowel in the *F*1–*F*2 space is larger than in other languages, with more vowels such as French and Norwegian (Crothers, 1978). This circumstance implies that there are larger ambiguous receptive areas across adjacent vowels (Ito et al., 2001; Ueda & Watanabe, 1987; Fig. 1). If five complex tones that are located less sparsely but symmetrically to the five-vowel pentagon in the *F*1–*F*2 space are presented in isolation, they can sound like a complete set of Japanese vowels pronounced ambiguously.

In the middle and last portions, the attenuation of the N1m in the pitch-formant sequence was not significant, although the attenuation of the N1m in the pitch sequence was significant. Shestakova et al. (2002) reported that the amplitudes of the N1m that are elicited by between-categorical speech sounds were significantly larger than those elicited by within-categorical speech sounds. Furthermore, they detected MMN using between-categorical speech sounds. In the present study, the pitch-formant sequence consisted of between-categorical speech sounds (*F*1, *F*2 variable), whereas the pitch sequence consisted of within-categorical speech sounds (*F*1, *F*2 constant). Therefore, our results are consistent with the findings of Shestakova et al.

In the formant sequence, we could not detect any significant difference in the neuromagnetic responses. This result might be consistent with the behavioral tests of the formant series in which the ratios of the correct answers on the familiarity tests were the lowest. The discrimination of the formants might be more difficult than the discrimination of the pitches. According to previous studies, the combination of melody and language (i.e., song) might facilitate language learning, especially in the first learning phase of language (Schön et al., 2008). Our findings are consistent with the results of these studies. On the other hand, the behavioral performance on the formant series was significantly above the chance level of 50%. Therefore, the formant sequences could also be learned, although we could not detect the learning effects from the neuromagnetic responses. Neuromagnetic evaluation might be less sensitive to the effects on statistical learning due to the methodological limitation of averaging compared with behavioral evaluation.

Another reason why statistical learning was less effective in the formant sequences compared with the pitch and pitch-formant sequences is that we cannot exclude the possibility that the participants dominantly depended on the pitch information during the statistical learning in the pitch and pitch-formant sequences. N1m varied considerably across the sequences with larger ampli-

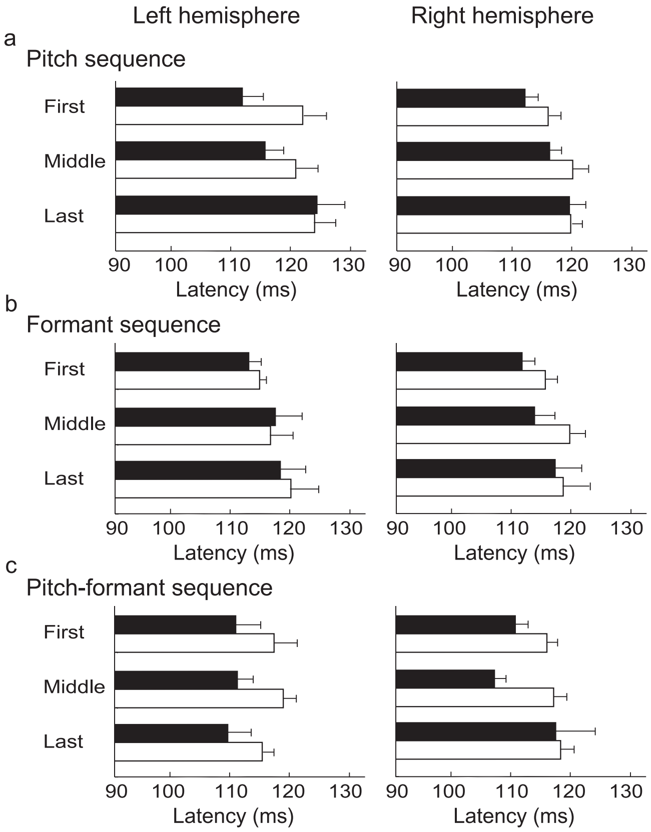


Fig. 7. The averaged peak latencies of the N1m responses ($N = 14$). The pitch (a), formant (b), and pitch-formant (c) sequences. The responses in the left and right hemispheres are located in the left and right sides, respectively. The closed bars represent the responses to the frequent tones, and the open bars represent the responses to the rare tones. The error bars indicate the standard error of the mean.

tudes for the pitch-formant sequence. $F0$ variability may recruit more neural activity than only $F1$ and $F2$ variability.

Many previous studies reported that the MMN could be used as an index for the differential processing of probable and improbable tones (or standards/oddballs) (Paraskevopoulos et al., 2012; Shestakova et al., 2002). In the present study, the differences in waveforms in the responses to rare and frequent tones are shown in Fig. 6. The temporal profiles of those difference waveforms were similar to the MMN, especially from the viewpoint of the peak latencies. The difference waveforms in our study might be interpreted as the MMN in the higher-order structure of the tone sequences. The neural responses to rare tones during statistical learning and the MMN might share the neural substrates for the processing of the improbable tones.

According to our knowledge, there are few studies that have investigated neurophysiological markers during statistical learning. A previous study detected statistical learning effects on P50 (Paraskevopoulos et al., 2012). However, in our study, few participants elicited P50, mainly because of the shallow rising of each tone onset. A previous study detected the statistical learning effects on N400 (Abla et al., 2008). The longer ISIs are necessary to record the N400. Because the ISI in the present study was 200 ms, it was difficult to detect the N400 in our study. Although Furl et al. (2011) found statistical learning effects on P2m as well as N1m in using pure tones, we could not detect the effects on P2m. This insignificance could result from the general notion that the P2m resists adaptation more than the N1m, and this tendency is more pronounced when the vowel sounds are used as repeated stimuli (Kuriki et al., 2007; Teismann et al., 2004).

The issue of domain generality versus domain specificity in the mechanisms of learning is an important topic for understanding how humans acquire language, music and other skills. Through investigations of how learning could occur, several studies have demonstrated that learners might use the same statistical learning mechanism for both linguistic (Peretz, Saffran, Schön, & Gosselin, 2012; Romberg & Saffran, 2010; Saffran, 2003b; Saffran, Loman, & Robertson, 2001) and musical stimuli (Furl et al., 2011; Saffran, 2003b; Saffran et al., 1999). Saffran and colleagues demonstrated that 8-month-old infants could recognize the difference between novel and familiar syllables of language within a two minute listening experience, which suggests that the infants were able to extract the information about sequential statistics because knowledge, such as music and language, must be acquired from a listening experience, especially in infants (Saffran et al., 1996). Furthermore, other reports suggest that statistical learning was not limited to auditory materials, such as music and language, displaying many analogous properties for visual pattern learning (Asaad, 1998) and visuo-motor sequence learning (Hunt & Aslin, 1998). Therefore, statistical learning has been conceived as a domain-general mechanism not only in auditory learning but also in learning in other modalities.

However, one question arises. Domain-general statistical learning does not function properly in the case of congenital amusia. In other words, individuals with congenital amusia fail to learn music but can learn language (Ayotte, Peretz, & Hyde, 2002), even if both of these constructs are organized around the same statistical properties (Peretz et al., 2012). Peretz et al. stated that the input and output of the statistical computation might be domain-specific, whereas the learning mechanism might be domain-general. This model was supported by the work of Omigie and Stewart (2011). Their results suggested that individuals with amusia lack confidence in their statistical learning ability, although they may statistically learn music. In other words, these authors suggested that amusia might be a disorder of awareness rather than of perception. Furthermore, in the present study, although the statistical learning was clearly confirmed from the MEG data, none of the participants could verbalize their statistical knowledge in detail. However, by using an abstract medium, such as a musical melody, they could correctly answer far above the chance level of 50% whether the tone series in the behavioral tests sounded familiar based on the tone sequences that had been presented in the measurement sessions. These findings may support the model of Peretz et al. (2012). Furthermore, in the present study, we also reveal the possibility of the statistical learning of language-like tone sequences (i.e., both the pitch and formant could be changed, such as in spoken language) as well as music-like tone sequences (i.e., only the pitch can be changed, such as the melody). This finding suggests that statistical learning might be a domain-general mechanism that is involved in both music and language.

In contrast, other studies on domain-specific learning mechanisms have suggested that similarities between language and music might be specific compared with any other auditory stimuli (Jackendoff & Lerdahl, 2006). Jackendoff and colleagues stated that the perception of both the music and spoken language are in part subject to the gestalt principle. A pitch that is a large interval away from the melody's surrounding context is perceptually segregated as an isolated pitch. Similarly, in spoken language, large frequency differences can be perceived distinctively. Furthermore, Hauser and Chomsky et al. have also claimed that many aspects of language structure cannot be described by finite state regularities. Therefore, they suggest that music and language cannot be acquired by statistical learning alone. In summary, although there are two conflicting claims (domain specificity versus domain generality) in studies on learning, there has been a recent consensus that learning depends on a combination of domain-specific and domain-general mechanisms. However, most theories of language acquisition have emphasized the critical role that is played by domain-specific structures over the role of domain-general structures (Hauser et al., 2002). Accordingly, the empirical studies and theoretical arguments are far from conclusive.

Earlier studies have revealed that the relative processing of an atonal melody that is not restricted to the diatonic scale is more difficult than the processing of a tonal melody (Dowling & Fujitani, 1971). This finding suggests that the relative processing of pitch could be acquired by a domain-specific mechanism for music (i.e., tonality). However, other studies have also reported that animals (Wright, 2007) as well as human infants (Trehub & Hannon, 2006) are capable of relative processing of pitch. Additionally, we clarified that the relative processing of auditory sequences can occur when the formant as well as the pitch was shifted. These results indicate that the auditory relative processing is available in domain-general statistical learning as well as domain-specific learning. In summary, our results suggest that the relative processing of auditory stimuli is an essential mechanism that is innate to humans. To reveal how statistically acquired knowledge could be appropriated, it is important to investigate higher-order statistical learning mechanisms other than relative processing.

5. Conclusions

This study demonstrated that the N1m response could be a marker for statistical learning. The pitch change may facilitate statistical learning in language and music. The relative processing of a spectral sequence may be a domain-general auditory mechanism that is innate to humans.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*, 952–964.

Archibald, L. M., & Joanisse, M. F. (2013). Domain-specific and domain-general constraints on word and sequence learning. *Memory & Cognition, 41*, 268–280.

Asaad, P. (1998). Statistical learning of sequential visual patterns. *Unpublished senior honors thesis*. New York: University of Rochester.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238–251.

Crothers, J. (1978). Typology and universals of vowel systems. In J. Greenberg (Ed.). *Universals of human language, phonology* (Vol. 2, pp. 93–152). Stanford, California: Stanford University Press.

Daikoku, T., Yatomi, Y., & Yumoto, M. (2014). Implicit and explicit statistical learning of tone sequences across spectral shifts. *Neuropsychologia, 63*, 194–204.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*, 524–531.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, England: Multilingual Matters.

Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *Neuroimage, 54*, 2267–2277.

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience, 2525*, 10494–10501.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1570–1582.

Hunt, R. H., & Aslin, R. N. (1998). Statistical learning of visuomotor sequences: Implicit acquisition of subpatterns. In *Proceedings of the twentieth annual conference of the cognitive science society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ito, M., Tsuchida, J., & Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception. *Journal of the Acoustical Society of America, 110*, 1141–1149.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*, 33–72.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*, 323–328.

Kewley-Port, D., & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America, 95*, 485–496.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology, 4*, e1000209. http://dx.doi.org/10.1371/journal.pcbi.1000209.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971–995.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., et al. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain, 123*, 2512–2518.

Kuriki, S., Ohta, K., & Koyama, S. (2007). Persistent responsiveness of long-latency auditory cortical activities in response to repeated stimuli of musical timbre and vowel sounds. *Cerebral Cortex, 17*, 2725–2732.

Markov, A. A. (1971). Extension of the limit theorems of probability theory to a sum of variables connected in a chain, reprinted in Appendix B of: R. Howard. D. *Dynamic Probabilistic Systems, volume 1: Markov Chains*. John Wiley and Sons.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544–2590.

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters, 265*, 119–122.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology, 14*, 481–487.

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

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia, 50*, 341–349.

Peretz, I., Saffran, J., Schön, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences, 1252*, 361–367.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175–184.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition, 98*, 1–11.

Poon, H., & Domingos, P. (2007). Joint inference in information extraction. In *Proceedings of the twenty-second national conference on artificial intelligence* (pp. 913–918). Vancouver, Canada: AAAI Press.

Poon, H., & Domingos, P. (2008). Joint unsupervised coreference resolution with Markov logic. In *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 650–659). Honolulu, Hawaii: Association for Computational Linguistics.

Rao, R. P. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport, 16*, 1843–1848.

Richardson, M., & Domingos, P. (2006). Markov logic networks. *Machine Learning, 62*(1–2), 107–136.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 906–914.

Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research, 248*, 48–59.

Saffran, J. R. (2003a). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science, 6*, 35–43.

Saffran, J. R. (2003b). Musical learning and language development. *Annals of the New York Academy of Sciences, 999*, 397–401.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Saffran, J. R., Loman, M. M., & Robertson, R. R. (2001). Infant long-term memory for music. *Annals of the New York Academy of Sciences, 930*, 397–400.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106*, 975–983.

Schönwiesner, M., Rübsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience, 22*, 1521–1528.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *Neuroreport, 13*, 1813–1816.

Singla, P., & Domingos, P. (2006). Entity resolution with markov logic. In *ICDM '06: Proceedings of the sixth international conference on data mining* (pp. 572–582). Washington, DC, USA: IEEE Computer Society.

Taulu, S., & Hari, R. (2009). Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses. *Human Brain Mapping, 30*, 1524–1534.

Teismann, I. K., Sörös, P., Manemann, E., Ross, B., Pantev, C., & Knecht, S. (2004). Responsiveness to repeated speech stimuli persists in left but not right auditory cortex. *Neuroreport, 15*, 1267–1270.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science, 7*, 289–296.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences, 930*, 1–16.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*, 73–99.

Ueda, Y., & Watanabe, A. (1987). Visible/tactile vowel information to be transmitted to the hearing impared. *The Journal of the Acoustic Society of Japan, 8*, 99–108.

Wright, A. A. (2007). An experimental analysis of memory processing. *Journal of the Experimental Analysis of Behavior, 88*, 405–433.

Yumoto, M., Matsuda, M., Itoh, K., Uno, A., Karino, S., & Saitoh, O. (2005). Auditory imagery mismatch negativity elicited in musicians. *Neuroreport, 16*, 1175–1178.