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/

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