Enhancement of speech-relevant auditory acuity in absolute pitch possessors

Nobuo Masataka*
Primate Research Institute, Kyoto University, Inuyama, Japan

INTRODUCTION
Absolute pitch (AP) is the ability to identify the frequency or musical name of a specific tone, or to identify a tone without comparing it with any objective reference tone. While AP has recently been shown to be associated with morphological changes and neurophysiological adaptations in the planum temporale, a cortical area in the brain involved in speech perception processes, no behavioral evidence of speech-relevant auditory acuity in any AP possessors has hitherto been reported. In order to seek such evidence, in the present study, 15 professional musicians with AP and 14 without AP, all of whom had acquired Japanese as their first language, were asked to identify isolated Japanese syllables as quickly as possible after these syllables were presented auditorily. When the mean latency to the syllable identification was compared, it was significantly shorter in AP possessors than in non-AP possessors whether the presented syllables were those used as Japanese labels representing the 7 tones constituting an octave or not. The latency to hear the stimuli per se did not differ according to whether the participants were AP possessors or not. The results indicate the possibility that possessing AP provides one with extraordinarily enhanced acuity to individual syllables per se as fundamental units of a segmented word in the speech stream.

Keywords: absolute pitch, syllable perception, music, planum temporale

MATERIALS AND METHODS
Participants
Fifteen professional musicians with AP (referred to below as “AP-possessors”: 3 males and 12 females; mean age = 23 years, SD = 3.3; mean practice years = 15.4, SD = 4.3; mean age practice began = 4.2, SD = 3.0), and 14 professional musicians without AP (referred to below as “non-AP possessors”: 4 males and 10 females; mean age = 23 years, SD = 3.8; mean practice age = 13.9, SD = 6.2, mean age practice began = 9.1, SD = 5.2) participated in the present study. All of them were born in Japan, had been educated in the Japanese formal education system from the beginning of the entrance into kindergarten at least until the completion of graduation from senior high school, and had acquired Japanese as their first language. None of them reported any hearing impairment.
Both AP and Non-AP possessors as participants were selected preliminarily with an in-house test which was conducted prior to the present experiment: they heard 108 pure sine wave tones, presented in pseudorandomized order, which ranged from A3 (tuning: A4 = 440 Hz) to A5, with each tone being presented three times. Each tone of the AP test had a duration of 1 s, with a 4-s interstimulus interval. During the intervals, the participants heard brown noise. They had to write down the tonal label immediately after hearing the accordant tone. The whole test unit and its components were created with Adobe Audition 1.5. The accuracy was evaluated by counting correct answers and the semitone errors were taken as incorrect to increase the discriminatory power. The participants were not asked to identify the adjacent octaves of the presented tones because for AP it is a most notable prerequisite to identify the correct chroma. In all, AP possessors were those whose accuracy scores were above 80% (mean = 85.8; SD = 6.6) whereas non-AP possessors were those whose scores were below 10% (mean = 7.2; SD = 3.9).

PROCEDURE
During the experiment as well as during the preliminary screening test for AP, each participant was seated in an attenuation chamber and wore a headphone. While a pure tone was presented using a notebook computer in a screening test, a syllable was chosen for a presented stimulus from a total of the 111 syllables that constitute the Japanese language. A total of 100 isolated syllables, each of which had a duration of 200 ms, were presented to a given participant consecutively with an interstimulus interval being 5 s of silence in a given presentation session, and in all, two such sessions were conducted for each participant. In the first session, the participant was asked to press a key which was located on a table near the participant as quickly as possible when identifying what syllable was the stimulus and to answer it orally (referred to below as “the identification session”). In the second session, the participant was asked to press the key as quickly as possible when hearing the presented sound (referred to below as “the hearing session”). Prior to the identification session, the participant was also instructed not to press the key before recognizing the presented syllable, and that was actually confirmed in each participant by an interview undertaken after the completion of the entire experiment.

As the presented stimuli in each session, the 111 syllables were operationally classified into two categories; seven were those that are used as Japanese tonal labels for seven musical notes constituting an octave, i.e., do (C), re (D), mi (E), fa (F), so (G), ra (A), si (B) (referred to below as “solfege syllables”), and the remaining 104 were those that are not used as note-names (referred to below as “non-solfege syllables”). The 100 isolated syllables presented to each participant in a given session comprised 50 solfege syllables and 50 non-solfege syllables. For each of the two categories, the presented stimuli were randomly chosen. Moreover, all of the stimuli were presented randomly to each participant regarding whether they were solfege syllables or non-solfege syllables.

As a behavioral measure, in both sessions, the interval between the onset of each stimulus presentation and the onset of the subsequent pressing of the button was used. The mean latency to the answer was computed for each participant in each of the two sessions separately with regard to solfege syllables and to non-solfege syllables.

RESULTS
Figure 1 shows the mean latency to press the button of AP possessors and of non-AP possessors in the identification session when the presented stimuli were solfege syllables as well as non-solfege syllables. Throughout the entire experiment, no identification errors were recorded and all the participants answered all the presented syllables correctly. Nonetheless, 2 (AP possessor versus non-AP possessor) × 2 (solfege syllable versus non-solfege syllable) analysis of variance (ANOVA) revealed a significant main effect [F(1,27) = 12.176, p = 0.002], and the mean latency to the presented stimulus was significantly shorter in the AP possessors than in non-AP possessors. The score was not significantly different whether the stimulus was a solfege syllable or a non-solfege syllable [F(1,27) = 0.012, p = 0.912]. Interaction between the two main factors was not significant, either [F(1,27) = 0.042, p = 0.839].

Figure 2 shows the mean latency to press the button of AP possessors and of non-AP possessors in the hearing session when the presented stimuli were solfege syllables as well as non-solfege syllables.
ANOVA revealed no significant main effects, and the average latency to the presented stimulus was not different between the AP possessors and the non-AP possessors \( [F(1,27) = 0.392, p = 0.536] \). The score was not different whether the stimulus was a solfege syllable or a non-solfege syllable \( [F(1,27) = 0.963, p = 0.334] \). Interaction between the factors was not significant, either \( [F(1,27) = 1.359, p = 0.254] \).

**DISCUSSION**

Previous research demonstrated that the basic auditory capability does not differ between AP possessors and non-AP possessors (Fujisaki and Kashino, 2002). The results of the hearing session in the present study are consistent with that conclusion. Nonetheless, recent neuroimaging studies have provided suggestive evidence for a strong influence of the pitch-processing expertise of AP possessors on their speech perception (Oechslin et al., 2010). The plausibility of this notion was tested in the present experiment, which actually presented suggestive evidence for such a link between musical expertise and speech information processing, and the results of the identification session in the present experiment revealed the fact that AP possessors are significantly superior in basic speech processing to non-AP possessors. Namely, AP possessors were able to identify a given isolated syllable chosen from their first language significantly more rapidly than non-AP possessors could, whether the syllable was one used as musical note or not.

Human infants are born with the predispositional capability to distinguish all the sounds in all of the world’s languages (Kuhl, 2000). By the end of their first year, however, they are on their way to perceiving particularly well the sounds that are important for their native languages (usually around 40 for a given language) whereas their capability to distinguish foreign speech sounds has decreased (Kuhl, 2003). Japanese infants, for example, initially perceive separate sounds for “r” and “l” (as in the words “road” and “load”) but lose the ability to hear this “foreign” distinction as they mature and become more adept at recognizing Japanese speech sounds.

Meanwhile, the first recognizable speech infants produce by themselves comprises a single word or what may appear to be a phrase, though at this stage, they are not aware that the words they produce have constituent elements. They do not understand the notion of word or lexical meaning, either. The question that then arises relates to the segmentation problem: how do children discover the structural components of the fluent speech stream without knowing the identity of the target elements? In fact, the infant’s task of learning its native language is a daunting one because, unlike written language, spoken language has no obvious markers that cover the structural components of the fluent speech stream with-
