# Absolute Pitch and Relative Pitch in Music Students in the East and the West: Implications for Aural-Skills Education

Ken'ichi Miyazaki
*Niigata University, Niigata, Japan*

Andrzej Rakowski
*Fryderyk Chopin University of Music, Warsaw, Poland*

Sylwia Makomaska
*University of Warsaw, Warsaw, Poland*

Cong Jiang
*Capital Normal University, Beijing, China*

Minoru Tsuzaki
*Kyoto City University of Arts, Kyoto, Japan*

Andrew J. Oxenham
*University of Minnesota*

Gregory Ellis
*University of Louisville*

Scott D. Lipscomb
*University of Cincinnati*

**Absolute pitch (AP)—an ability to identify** an isolated pitch without musical context—is commonly believed to be a valuable ability for musicians. However, relative pitch (RP)—an ability to perceive pitch relations—is more important in most musical contexts. In this study, music students in East Asian and Western countries (Japan, China, Poland, Germany, and USA) were tested on AP and RP abilities. In the AP test, 60 single tones were presented in a quasirandom order over a five-octave range. In the RP test, ascending musical intervals from 1 to 11 semitones were presented in four different keys. Participants wrote down note names in the AP test and scale-degree names or musical interval names in the RP test. The conservatory-level Japanese students showed the highest AP performance and more than half of them were classified as accurate AP possessors, but only 10% were classified as accurate RP possessors. In contrast, only a small percentage of participants from Poland, Germany, and the USA were identified as accurate AP possessors, whereas many more were accurate RP possessors. Participants from China were typically intermediate on both measures. These noticeable contrasts between AP and RP performance in different countries suggest influences of the underlying socio-cultural conditions, presumably relating to music education. Given the importance of RP in music, the results suggest that more emphasis should be place on RP training, particularly in East Asian countries.

*Received: December 10, 2017, accepted May 11, 2018.*

**Key words:** absolute pitch, relative pitch, cross-cultural, tonality, music education

**Absolute pitch (AP)—an ability to name** or produce a note of a given pitch in the absence of a reference note (Deutsch, 2013)—has attracted much attention from scientists and the general public because it is understood to be extremely rare and is often regarded as a sign of exceptional musical ability (Deutsch, 2002; Levitin & Rogers, 2005; Takeuchi & Hulse, 1993; Ward, 1999). The rarity of AP has been repeatedly referred to in the scientific literature since Bachem's seminal work, in which he estimated the incidence of AP in the general population to be less than 1 in 10,000 (Bachem, 1955), based on the number of AP possessors he identified in samples of professional musicians and music students in the Chicago area (Bachem, 1937, 1940). Although this estimate is misleading (see Miyazaki, Makomaska, & Rakowski, 2012, for more detail), AP may indeed be rare in the general population (i.e., including both musicians and nonmusicians). However, more meaningful is the prevalence of AP among people with musical experience, since AP is a music-related ability that depends on certain types of music training.

Several large-scale studies have tested music students and have provided estimates of the prevalence of AP among musicians. Deutsch, Henthorn, Marvin, and Xu (2006) tested students of the Eastman School of Music (Rochester, New York) and found 5 AP possessors out of 105 students (approximately 5%) with an operationally defined criterion for AP set at more than 85% correct responses in their AP test. Miyazaki et al.

*Music Perception*, volume 36, issue 2, pp. 135–155, issn 0730-7829, electronic issn 1533-8312.  DOI: https://doi.org/10.1525/mp.2018.36.2.135

(2012) tested 250 students of the Chopin University of Music in Warsaw, Poland, and found 10% of them exceeded the criterion for AP used by Deutsch et al. (2006). Furthermore, Leite, Mota-Rolim, and Queiroz (2016) showed that the percentage of music students with AP was 4% in Brazil (a criterion of 85% correct), and Chavarria-Soleya (2016) showed it was 13% (a 90% correct criterion) in Costa Rica. Based on these empirical results, the percentage of AP possessors among music students in America and Europe might be estimated at around 10%, a percentage that varies depending on the sampled population, the method of the AP test, and the AP threshold used. It is not surprising that the AP prevalence is much higher among music students than in the general population, especially given the inherent bias in the experimental response method (i.e., participants must understand and be able to respond using musical pitch labels). However, the fact remains that AP possessors are a small minority of the population even among musicians in the Western countries.

In contrast to this, a few reports have demonstrated that AP is more prevalent among music students in East Asia than in North America and Europe. Deutsch et al. (2006) carried out an on-site AP test at the Central Conservatory of Music (Beijing, China) and reported that 55% of the participants met a criterion of at least 85% correct (calculated from the data shown in a figure in Deutsch et al., 2006). Deutsch, Li, and Shen (2013) also conducted the same AP test at the Shanghai Conservatory of Music in China and found that the participants showed a similarly high percentage of accurate AP possessors. Furthermore, Miyazaki et al. (2012) examined the AP ability of Japanese students majoring in music education in Niigata University in Japan using a similar AP test and showed that the percentage of accurate AP defined by a criterion of at least 95% correct was approximately 30%, which corresponded to nearly 45% using Deutsch et al.'s AP criterion of at least 85% correct. The AP scores were continuously distributed from 100% accurate to chance performance, with many participants revealing a moderate level of AP. These results indicate that, in East Asia (at least in China and Japan), a large proportion of music students have AP, varying in degree of accuracy, contradicting the prevailing view that only a limited number of musicians have AP.

Significant interest in AP also exists due to the prevailing notion that it is a valuable ability for musicians and increases the likelihood of attaining outstanding musicianship. Arguments for this notion often refer to anecdotal reports about well-known musicians who reportedly had AP, such as Mozart, Beethoven, Toscanini, and Boulez (Deutsch, 2002). However, such arguments must be called into question. Some of these distinguished musicians might have had AP, but it is difficult—if not impossible—to confirm whether or not musicians from past eras had AP, as this can only be surmised based on biographical or anecdotal information. More important, these arguments commit a fallacy of logic, an illusory correlation between high levels of musicianship and AP possession, focusing only on musicians supposedly having AP and ignoring equally distinguished—and certainly many more—musicians without AP.

Another type of argument for the notion favoring AP draws on observations of the astonishing achievement of music students with AP in various musical activities, particularly in aural-skills tasks such as music dictation and sight-singing. These observations are not only frequently found in informal anecdotal reports both by music teachers and by musicians but are also documented experimentally in the scientific literature. Dooley and Deutsch (2010) administered an AP test and a melodic dictation test to music students and found a strong positive correlation between the scores of the two tests. It is not surprising that AP possessors achieve higher scores on this task, because it is quite easy for them to name heard tones of a melody instantaneously and to write down those individual notes on a musical staff. AP would function as a useful tool for such a task. Furthermore, AP possessors would presumably enjoy an advantage in sight-singing (i.e., the ability to sing a musical excerpt that has never been seen before), although there has been no experimental evidence yet provided to support this conclusion. The definition of AP often specifies the ability to produce (sing) pitches of designated notes as an active aspect of AP; therefore, AP would directly enhance an individual's ability to perform a sight-singing task.

In a subsequent study, Dooley and Deutsch (2011) administered an interval-naming test. In one condition, participants heard a V7-I chord sequence (a cadence that clearly establishes a tonal context) and an interval formed by a sequence of two tones, the first tone of which was the tonic of the established key. The stimulus sequences were presented in different keys. The participants were asked to write down the name of each interval after they heard it. The researchers suggested that this was a stronger test of the relevance of AP to performance on musical tasks that could be unrelated to AP possession, because performance in this test depended only on perceiving the relationship between the two pitches comprising the interval. The results showed that the AP group performed significantly better than the non-AP group. Indeed, perception of musical intervals in different key contexts is a fundamental skill that is indispensable to musicians. Superficially, these results seem to support the hypothesis that AP possessors have an advantage in a musically relevant task that requires only relative pitch. However, the interval-naming test used by Dooley and Deutsch (2011) would not provide convincing evidence for the musical relevance of AP unless the use of the AP strategy were suppressed, resulting in the same problem identified previously in relation to the music dictation test. At least some participants with accurate AP would find it much easier to identify the AP names of the two tones and to figure out an interval name by counting the number of scale steps between the two tones than to identify an interval solely relying on relative pitch.

Putting these methodological issues aside, just because music students with AP perform exceptionally well on a music dictation or sight-singing test does not mean they are excellent in aural skills and enjoy a high degree of musicianship. No matter how useful AP seems to be as a tool for musicians, it does not include musically meaningful processing but simply works like computer software, translating incoming sounds into a sequence of MIDI note codes or converting the notated pitches into sounds. Aural skills (solfège skills or musical literacy) entail musical listening and reading—essential abilities for musicians—the essence of which is representing music in the mind based on incoming sounds (musical listening) or by internally hearing music when viewing musical notation (sight reading or inner hearing) (Karpinski, 2000, pp. 1–8). The problem is that the internal representation of incoming music cannot be observed directly but must be hypothesized from the observed performance on a dictation or sight-singing test, so it is impossible to know with complete confidence to what extent participants successfully form such representations. It would be difficult to ascertain whether the participants drew on AP as a short-cut strategy to tackle the task; therefore, the test could not affirm with certainty whether some of the participants exploited AP, resulting in a higher score on the task.

In contrast to Dooley and Deutsch (2010, 2011), in which the authors claim demonstrated advantages of AP in the melodic dictation task and in the music-interval identification task, one of the present authors (Miyazaki, 1993, 1995) has provided contradictory results that demonstrate musical disadvantages of AP. Using interval identification, Miyazaki conducted a series of experiments in which different musical intervals were presented in varying tonal contexts (C major, F♯ major, and E major shifted downward by a quarter-tone). The music students having AP showed significantly poorer performance in identifying the intervals than those who did not possess AP in the key conditions other than the C major condition. This may be related to the key-color effect—white-key notes are easier than black-key notes (Deutsch, 2013, for review). Moreover, the disadvantage of AP possession was also demonstrated in experiments on melody recognition (Miyazaki, 2004; Miyazaki & Rakowski, 2002), in which participants with AP performed significantly worse in discriminating between a standard melody in sound or visually presented in the form of notation in the key of C major and a comparison aural melody transposed to a different key.

These conflicting results regarding the musical advantage or disadvantage of AP might be explained by the differences in sampled populations that varied widely in the music education received and sociocultural circumstances in which they obtained this education. In an attempt to examine these cross-cultural differences, we designed the present study, comparing AP and relative pitch ability among music students in different countries including both Eastern and Western cultures. AP is a relatively simple ability that can be defined clearly, as shown at the beginning of this paper. We used a conventional AP test devised according to the standard definition of AP, in which individual tones in piano timbre comprising the chromatic scale were presented in a random order and participants were required to name each tone. On the other hand, relative pitch cannot be defined in such a straightforward way, particularly when trying to define relative pitch in a musically meaningful way. Relative pitch is often regarded as an ability to perceive musical intervals. However, musical intervals simply stand for pitch distances counted as the number of semitones between two notes and can be located at any pitch levels independent of the pitch context. Rather, in tonal music, locations of individual tones within a tonal context (musical scale) are more important than the simple pitch distances between a pair of tones. Melodies and harmony are comprised of a combination of different scale-degree tones, each having a specific tonal function. In fact, all tonal music theories and cognitive theories of the musical pitch hierarchy deal with relative pitch phenomena in this sense, with little or no consideration of AP (cf. Krumhansl, 1990; Krumhansl & Cuddy, 2010; Lerdahl, 2001, 2015; Lerdahl & Krumhansl, 2007). Hence, we use hereafter the term relative pitch (RP) to mean the scale-degrees and tonal functions in the tonal pitch context, not pitch distances.

We conducted an RP test, closely similar to that used by Dooley and Deutsch (2011), in which each trial consisted of an authentic cadence (a harmonic sequence proceeding from the dominant seventh chord to the tonic) to establish a tonal context, followed by a pair of single tones (the tonic tone of the established key and a target tone). The participants' task was to identify the scale-degree of the target tone in the established key.

The rationale for our study was as follows: AP is, by definition, an ability to identify the pitches of isolated tones without reference to any musical pitch context. Due to the fact that music is constructed on pitch relations within a tonal context, not on AP, it follows that AP is of minimal use in the cognitive processing of music, and may be less valuable to music students than often presumed. This statement might seem counterintuitive in that it is contrary to the commonly accepted notion favoring AP as a musically valuable ability. On the other hand, RP, as defined in the present study, is an essentially important ability for musicians; that is, an ability to perceive the scale-degrees and the tonal functions of musical tones within an established tonal context, a completely different ability from AP. Therefore, we propose that there is no significant correlation between AP and RP, as a weak hypothesis. Furthermore, AP processing can sometimes conflict with RP processing and may cause cognitive interference in musical activities that require RP processing. We therefore propose an even more counterintuitive notion as a strong hypothesis that AP is inversely related to RP, assuming a significant negative correlation between AP and RP. We have already reported experimental results obtained in a laboratory setting supporting this hypothesis (Miyazaki, 1993, 1995; Miyazaki & Rakowski, 2002).

The present study had two main objectives: one was to clarify the prevalence of AP in music students and its difference among East Asian and Western countries; another objective—musically more important—was to obtain information on the RP skills in the same populations to determine whether AP skills were associated positively or negatively with RP skills. We tested music students in Japan, China, Poland, Germany, and the USA to assess empirically their AP and RP abilities, and examined the relationship between the AP and RP performance from a cross-cultural perspective. To our knowledge, the present study is the first attempt to examine AP and RP abilities cross-culturally and to compare their relationship among music students from Eastern and Western cultures.

## Method

### PARTICIPANTS

In total, 1,016 undergraduate students at eight music institutions in five countries (Japan, China, Poland, Germany, and USA) participated. They were grouped according to their affiliations as shown in Table 1. The participants varied, to some extent, in the level of music training they had received depending on the institutions with which they were associated. The participants from JP-KCUA, CN-CCOM, CN-SCOM, and PL-UMFC had received conservatory-level music training for professional musicians or experts in musical specialties, while the other groups were relatively mixed: the JP-NGTU and CN-CNU groups majored in music education, the DE-MLU group majored in musicology, and the US-UMN group was from a music school at a public university that included music education, musicology, and music performance majors. There were a small number of foreign students (1 at CN-CCOM, 10 at PL-UMFC, 6 at US-UMN), who were excluded from data analysis on the basis of their native language and the place where they had received their principal music training, since the primary purpose of the present study

**TABLE 1.** *Participants Grouped by their Affiliations, their Age, and Onset Age of Music Training*

| Affiliation | Country | Number of participants | Average years of age (*SD*) | Average onset age of music training (*SD*) |
|---|---|---|---|---|
| Department of Music, Kyoto City University of Arts (JP-KCUA)* | Japan | 261 | 20.3 (1.4) | 5.3 (3.1) |
| Music Program of Faculty of Education, Niigata University (JP-NGTU) | Japan | 143 | 20.2 (2.2) | 5.0 (2.4) |
| Central Conservatory of Music, Beijing (CN-CCOM)* | China | 62 | 20.4 (1.5) | 6.7 (2.8) |
| Shanghai Conservatory of Music (CN-SCOM)* | China | 103 | 19.4 (1.0) | 7.4 (3.6) |
| School of Music, Capital Normal University, Beijing (CN-CNU) | China | 94 | 20.3 (0.6) | 6.8 (3.0) |
| Fryderyk Chopin University of Music, Warsaw (PL-UMFC)* | Poland | 117 | 21.5 (2.4) | 7.8 (3.1) |
| Institute of Musicology, Martin-Luther University, Halle (DE-MLU) | Germany | 57 | 22.3 (3.5) | 8.5 (4.2) |
| School of Music, University of Minnesota, Minneapolis (US-UMN) | USA | 162 | 20.5 (4.4) | 8.2 (3.3) |

* Conservatory-level institutions. See the Participants section for detail.

was to compare the AP and RP performance among groups of music students with different cultural backgrounds. Furthermore, there were 19 CN-SCOM students and 30 CN-CNU students majoring in Chinese traditional music. Most of them had started their music training in Chinese traditional instruments and had limited experience in Western music. Therefore, they were excluded from the calculation of the overall AP and RP scores of the institutions.

Responses to a questionnaire given to the participants showed that their average age at the time of testing ranged from 19.4 to 22.3 years and was significantly different between groups, $F(7, 975) = 10.42$, $p = .001$, $\eta_p^2 = .07$. Post hoc multiple comparisons revealed that the Japanese and Chinese groups were significantly younger than the Polish and German groups (Tukey HSD: $p < .05$). The average age of music training onset ranged from 5.0 to 8.5 years and was also significantly different between groups, $F(7, 979) = 23.30$, $p = .001$, $\eta_p^2 = .14$. The age of music training onset for the Japanese groups (JP-NGTU and JP-KCUA) was significantly earlier than other groups, and those of US-UMN and DE-MLU were significantly later than those of the Japanese and Chinese groups (Tukey HSD: $p < .05$).

### PROCEDURE

Participants were tested in a group setting in a classroom with exactly the same instructions (appropriately translated) and following identical procedures across all the groups. First, participants filled out a questionnaire, providing their native languages and basic information about their musical background. Then, they were provided with written instructions for the AP and RP tests and heard the instructions read out loud by a proctor (in most cases, one of the authors) before beginning each of the test sessions. The study was approved by the Ethical Committee of Niigata University and by the relevant local review boards.

*Absolute pitch (AP) test.* The AP test consisted of 60 isolated tones of the chromatic scale over five octaves, the fundamental frequencies of which ranged from C2 (65.4 Hz) to B6 (1975.5 Hz), tuned according to equal temperament with the standard A4 (440 Hz). The test tones were sampled piano sounds generated using a high-quality sound library (Vienna Concert Grand, Native Instruments) and edited using sound-editing software (Amadeus Pro, HairerSoft). They were presented in a quasirandom order with the restriction that successive tones were always different pitch classes separated by 14 or more semitones to discourage the use of relative pitch. Each individual tone was approximately 1s in length, and there was a 3-s interval between the onsets of successive tones. The sequence of the test tones was recorded onto a compact disc (CD) and played back on a CD player or from the computer's audio output using an audio system appropriate for the given classroom.

Preceding the experimental procedure, there were 10 practice trials, in which 10 tones randomly selected in the same way as the test tones were presented; after a short break, the test session consisting of 60 test tones followed. The AP test session, including practice trials, lasted approximately 4 min. After each test tone was presented, participants recorded what they perceived to be the pitch class label for the test tone, ignoring its octave position, in a numbered space on an answer sheet. For pitch class naming, participants were allowed to use their most familiar labels: musical pitch names (e.g., C, C♯, D, etc.) or solfège syllables in the fixed-*do* system (e.g., do, do♯, re, etc., where "do" is equivalent to "C"). If uncertain about the correct answer, participants were encouraged to guess the pitch name. They were cautioned against relying on vocal cues and relative pitch.

*Relative pitch (RP) test.* The second test session was the RP test. The test sequence was constructed using the same sampled piano tones as in the AP test. In each trial, a sequence of a pair of chords forming an authentic cadence (the dominant seventh and the tonic chords) and a pair of successive tones was presented. The first tone of the tone pair was always the tonic of the key and the second tone (1- to 11-semitones higher than the tonic) was the target tone for which RP was to be identified. The chords and tone pairs changed from trial to trial between 4 different keys: C major, A♭ major, F♯ major, and E major, though for the last (E major) all the tones were transposed up by a quarter-tone compared to standard pitch (hereafter called E+ major). The duration of each chord and each tone was 1.0 s and 0.5 s, respectively; there was a silent interval of 0.5 s between the second chord and the first tone. The time interval between the onset of the target tone and the onset of the first chord of the next trial was 3.5 s, within which participants were required to record the responses in the numbered space on an answer sheet. Participants were instructed to identify the RP (scale-degree) of the target tone in the established key with the penultimate tone as the tonic. The participants were allowed to respond using the naming convention of labeling RP most familiar to them, including movable sol-fa names (mi, sol, ti flat, etc.), scale degree numbers (1-7 with + or -), musical interval names from the tonic (major 3rd [M3], perfect 5th [P5], minor 7th [m7], etc.).

There were 8 practice trials and, following a short break, 44 test trials. The practice and test trials lasted approximately 6 min.

## Results

### AP TEST PERFORMANCE

Performance on the AP test and the RP test will be expressed hereafter in proportion correct scores (1.0 for perfect performance). Figure 1 represents the percentages of participants that achieved certain accuracy levels from each institution. Performance is clearly different among the eight groups. The Japanese participants excelled in their AP accuracy; most remarkably, in JP-KCUA, one third of the participants had perfect AP, and more than half of them demonstrated accurate AP performance (defined as the AP score of 0.9 or above), whereas only a tiny minority were at the non-AP level (defined as the AP score less than 0.5). Although not comparable to the JP-KCUA students, the JP-NGTU students also performed fairly well, with approximately one-sixth of them perfect, one third of them at the accurate level, and only one-third of them at the non-AP level. Performance of the Chinese participants was lower than that of the Japanese participants. At CN-CCOM and CN-SCOM, performance was approximately comparable; roughly one-fourth of the participants were at the accurate level and about one half at the non-AP level. In a striking contrast, the Western students showed significantly lower performance than the Japanese and Chinese students on the AP test; students having accurate AP accounted for only 12% at PL-UMFC, and there were almost no AP possessors at DE-MLU and US-UMN.

Table 2 shows the median ($Q_2$) and lower/upper quartile ($Q_1/Q_3$) values of the AP scores for each

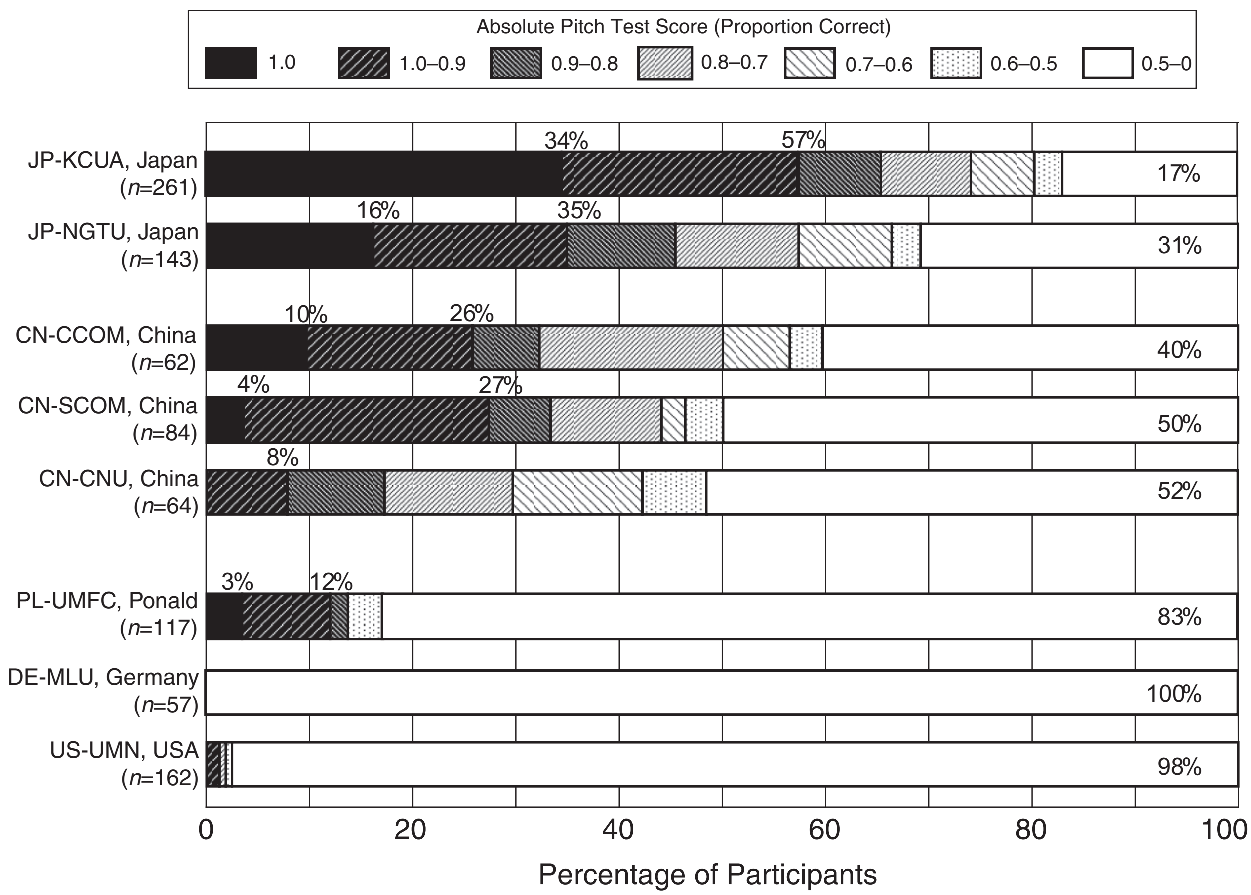



**FIGURE 1.** Percentages of the participants classified according to their AP scores for each group. JP-KCUA: Kyoto City University of Arts, JP-NGTU: Niigata University, CN-CCOM: Central Conservatory of Music (Beijing), CN-SCOM: Shanghai Conservatory of Music, CN-CNU: Capital Normal University (Beijing), PL-UMFC: Frederic Chopin University of Music, DE-MLU: Martin Luther University (Halle), US-UMN: University of Minnesota. Participants majoring in Chinese traditional instruments (19 CN-SCOM students and 30 CN-CNU students) were excluded. Numbers above each column indicate percentages of participants who achieved a perfect score (1.0) and an accurate score (0.9 or above).

**TABLE 2.** *Median and Lower/Upper Quartile Values of the AP Scores (Proportion Correct) with the Summary of the Multiple Comparisons Among Groups*

| | Median | | | Homogeneous Subsets | | | | |
|---|---|---|---|---|---|---|---|---|
| | $Q_1$ | $Q_2$ | $Q_3$ | | | | | |
| JP-KCUA (Japan) | 0.683 | 0.933 | 1.000 | a | | | | |
| JP-NGTU (Japan) | 0.425 | 0.750 | 0.967 | | b | | | |
| CN-CCOM (China) | 0.183 | 0.683 | 0.896 | | b | c | | |
| CN-SCOM (China) | 0.167 | 0.508 | 0.917 | | | c | | |
| CN-CNU (China) | 0.150 | 0.450 | 0.725 | | | c | | |
| PL-UMFC (Poland) | 0.050 | 0.150 | 0.333 | | | | d | |
| DE-MLU (Germany) | 0.050 | 0.083 | 0.100 | | | | | e |
| US-UMN (USA) | 0.050 | 0.083 | 0.117 | | | | | e |

Note: Participants majoring Chinese traditional music at CN-SCOM and CN-CNU were excluded from the analysis.

group. We evaluated between-group differences in test performance statistically using a nonparametric Kruskal-Wallis test, because the distributions of the AP scores of some groups were strongly skewed toward the lower or higher end. The analysis indicated that there were large and significant differences between the groups, $\chi^2 = 490.13$, $df = 7$, $p < .001$, $\eta^2 = .52$. Subsequently, we conducted follow-up multiple comparisons using the Mann-Whitney rank sum test with the Bonferroni correction ($p < .05$). The rightmost column of Table 2 provides a summary of the analyses, indicating separate subsets that include groups with scores not significantly different from each other. The performance of JP-KCUA was significantly higher than all other groups ($r = .29$–$.80$). The differences in performance among JP-NGTU and the Chinese groups (CN-CCOM, CN-SCOM, and CN-CNU) were small ($r = .12$–$.19$) and not significant. The Western groups (PL-UMFC, US-UMN, and DE-MLU) were in the lowest subsets with PL-UMFC significantly higher than US-UMN and DE-MLU ($r =$ .28 and .27, respectively).

Comparisons of the overall performance differences among these groups might be questioned, because the participants in each group were not comparable to each other in their music specialties, and AP performance may depend on the type of music training they had received. Therefore, we divided the participants of each of the conservatory-level groups (JP-KCUA, CN-CCOM, CN-SCOM, and PL-UMFC) into subgroups according to their specialties, and carried out a more detailed analysis. Figure 2 displays the percentages of participants associated with different levels of performance for the subgroups of different specialties in each institution. All subgroups at JP-KCUA demonstrated a high level of AP performance. Performance of the participants majoring in piano was exceedingly high; all of the piano majors had AP, remarkably, 93% of them at the accurate level. The strings subgroup and the composition/musicology subgroup also evinced high AP performance with 75% and 72% of them at the accurate AP level, respectively. Compared to these three subgroups, the AP performance of the subgroups of wind instruments and voice was much lower, but the percentages of accurate AP were still fairly high (31% and 28%, respectively) in comparison to the other groups. In contrast to JP-KCUA, the Chinese groups showed a different pattern of results in which accurate AP was observed only in particular subgroups; accurate AP was 50% of the subgroup of composition and conducting majors in CN-CCOM and 63% of the subgroup of piano majors in CN-SCOM. The proportion of accurate AP was small or zero in other subgroups of the Chinese participants. At PL-UMFC, the proportion of accurate AP was generally small (6–12%), except for the subgroup of sound engineering majors in which accurate AP accounted for 30% of participants.

We also assessed the relationship between AP performance and age of onset of music training. We divided the participants of each group into subgroups according to the self-reported age at which they began musical study and compared the AP performance among the subgroups. Figure 3A displays the average AP scores of the subgroups as a function of age of onset of music training. In general, the average AP scores of all groups decreased as the participants began their music training at later ages. Most notably, all subgroups of JP-KCUA showed fairly high AP scores; average AP scores of participants who began music training by three to seven years of age were 0.75 or above and the average AP score of the subgroup of eight and later onset age was still as high as 0.6. Figure 3B displays the percentage of accurate AP participants (AP score of 0.9 or greater) as a function of age of onset of music training. The percentages of accurate AP in all groups decreased for those who began music training later. Here also, JP-KCUA is notable in that the percentages of accurate AP participants were comparatively high for all subgroups, although the percentage clearly decreased as the onset age was later. Particularly interesting for JP-KCUA participants is that accurate AP occupied about 70% for the subgroups of four and earlier onset age, 50% for the subgroups of five to seven onset age, and still 30% for the subgroup of eight or later onset age. Compared with JP-KCUA, the percentages of accurate AP of other groups were generally smaller and almost negligible for the groups of seven or later onset age.

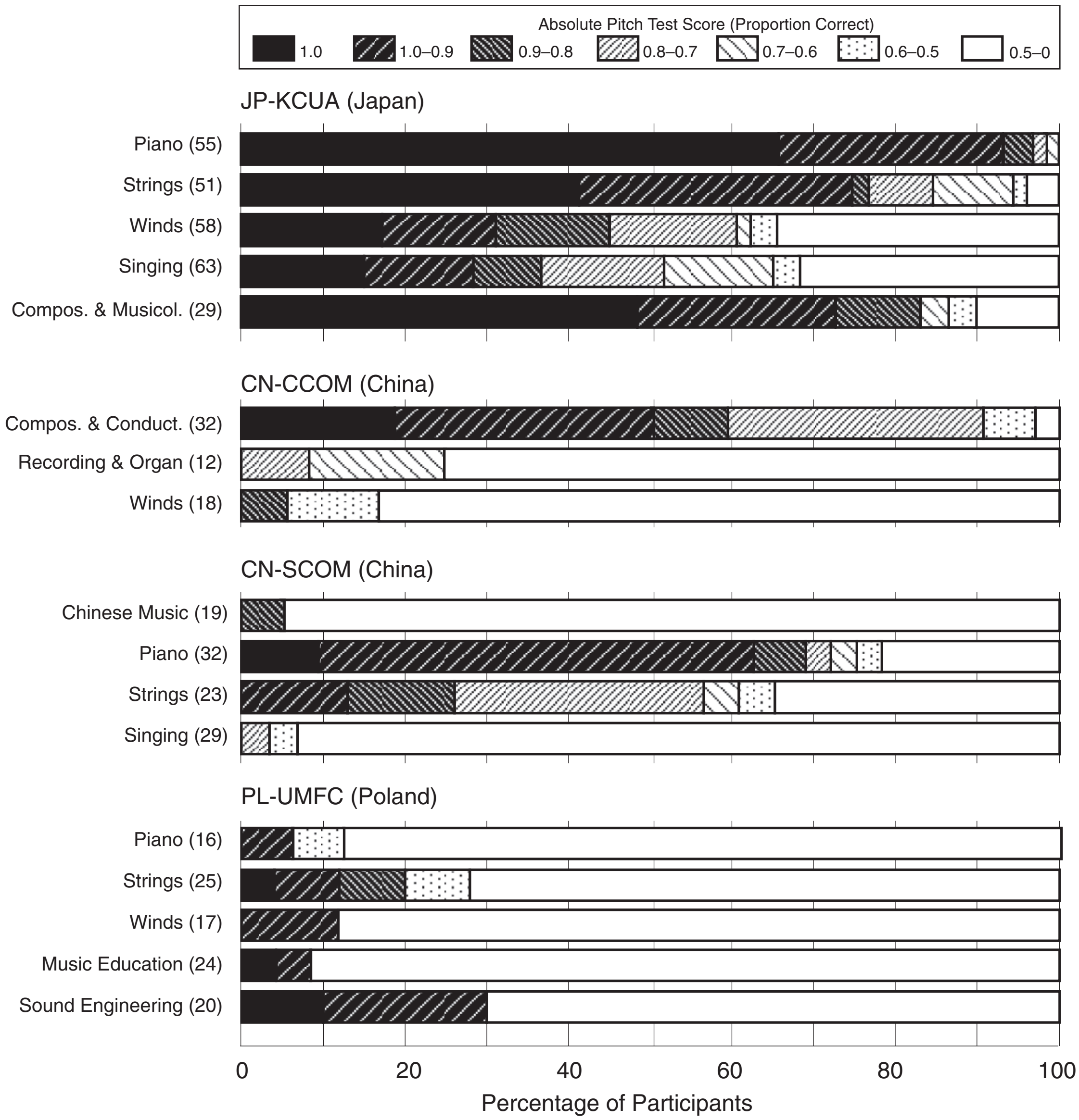
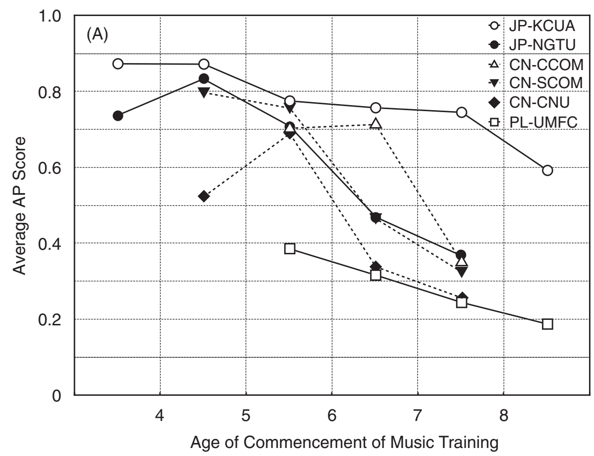



**FIGURE 2.** Percentages of the participants classified according to their AP scores for different specialties in each conservatory-level group.

RP TEST PERFORMANCE

In calculating the RP score, we excluded the C major condition, since the RP test would be virtually equivalent to the AP test for AP listeners in the C major condition in which they had only to answer the fixed-*do* sol-fa name of the target tone instead of the movable-*do* name demanded in the task, resulting in unfairly enhanced performance.

Results of the RP test are summarized in Figure 4. The overall RP performance exhibited remarkable differences between groups, as was the case for the AP performance, but the pattern of results was almost opposite in comparison to AP results. The Japanese music students, most of whom had AP, showed poor scores in the RP test; the percentage of accurate RP (0.9 or higher scores) was only 8% in JP-KCUA and 3% in JP-NGTU, with about one-third of the JP-KCUA students, and well over half of the JP-NGTU students, scoring 0.5 or less. In contrast, students in the Western countries showed much higher scores; remarkably, 72% of the PL-UMFC students scored higher than 0.9 on the RP task and very few scored 0.5 or lower. Performance of the Chinese groups was again in between the Japanese groups and the Western groups.




| | | | | | | |
|---|---|---|---|---|---|---|
| JP-KCUA | 70 | 63 | 53 | 32 | 10 | 33 |
| JP-NGTU | 38 | 30 | 34 | 22 | 18 | |
| CN-CCOM | | | 27 | 13 | 21 | |
| CN-SCOM | | 16 | 17 | 17 | 31 | |
| CN-CNU | | 12 | 18 | 16 | 17 | |
| PL-UMFC | | | 19 | 25 | 30 | 41 |

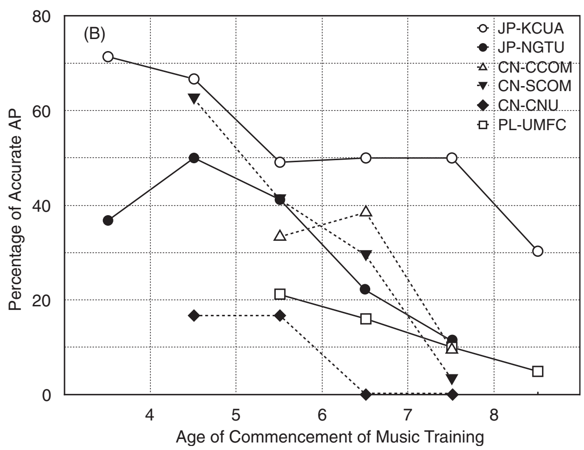



**FIGURE 3.** (A) Average AP scores and (B) percentages of accurate AP participants whose AP scores were 0.9 or higher for each group, as a function of age of onset of music training. Participants majoring in Chinese traditional instruments at CN-SCOM and CN-CNU were excluded. The data points at leftmost and rightmost ends of each line include the earlier and later years, respectively. Each row of numbers between the panels represents the numbers of participants of training onset age subgroups for each group.

There were some participants in the Japanese group (JP-KCUA and JP-NGTU) and the Chinese group (CN-CCOM and CN-CNU) who responded incorrectly using two pitch names in almost all trials instead of the required RP name of the target tone, despite the instructions given prior to the test, stating explicitly that the participants should write down the RP name (the movable-*do* sol-fa name or the interval name from the tonic). We designated this response as the AP response and displayed the participants who made the AP responses in most of the trials as the rightmost portion of each bar in Figure 4. The AP responses were essentially misguided responses, which ought to have been excluded from data analyses as response failures, but were counted here as a noteworthy response type indicating that some AP participants did not understand the instructions given and followed their customary practice relying on AP. This group accounted for about 5 to 10% in the Japanese and Chinese groups. Actually, the Japanese participants received an additional follow-up explanation telling them specifically not to answer with two pitch names but to use RP names, as instructed, because, at first, we regarded the AP responses as failures. There would likely have been more Japanese participants who provided these AP responses, if they had not received these additional, explicit instructions.

Table 3 shows the median and lower/upper quartile values of the RP score for each group. The nonparametric Kruskal-Wallis test indicated that the difference in the RP test score among groups was statistically significant, $\chi^2 = 291.56$, $df = 7$, $p < .0001$, $\eta^2 = .32$. The result of multiple comparisons using the Mann-Whitney rank sum test with the Bonferroni correction for $p < .05$ is shown in the right-most column of Table 3. The score of PL-UMFC was significantly higher than all other groups ($r = .34$–$.76$); CN-CCOM, CN-SCOM, and US-UMN formed the next higher group subset for which scores were similar and not significantly different from one another statistically ($r = .04$–$.12$), followed by the group subset of JP-KCUA, CN-CNU and DE-MLU with similar scores ($r = .04$–$.09$); the score of JP-NGTU was significantly lower than all other groups ($r = .25$–$.76$).

For further analysis, we divided the participants of each group into subgroups by their music specialties; JP-NGTU, CN-CNU, and DE-MLU were excluded from the analysis because the specialties of the participants of these groups were not suitably equivalent to allow meaningful subgroup comparisons. Results of this subgroup comparison are represented in Figure 5. The performance on the RP test varied among subgroups of different specialties in some institutions. In JP-KCUA, where the overall RP performance was the poorest among the conservatory-level groups in spite of achieving the highest AP performance, the students majoring in piano and composition/musicology showed relatively higher performance (at best 16% and 21%, respectively, at the accurate level) compared to the subgroups of other specialties for which performance

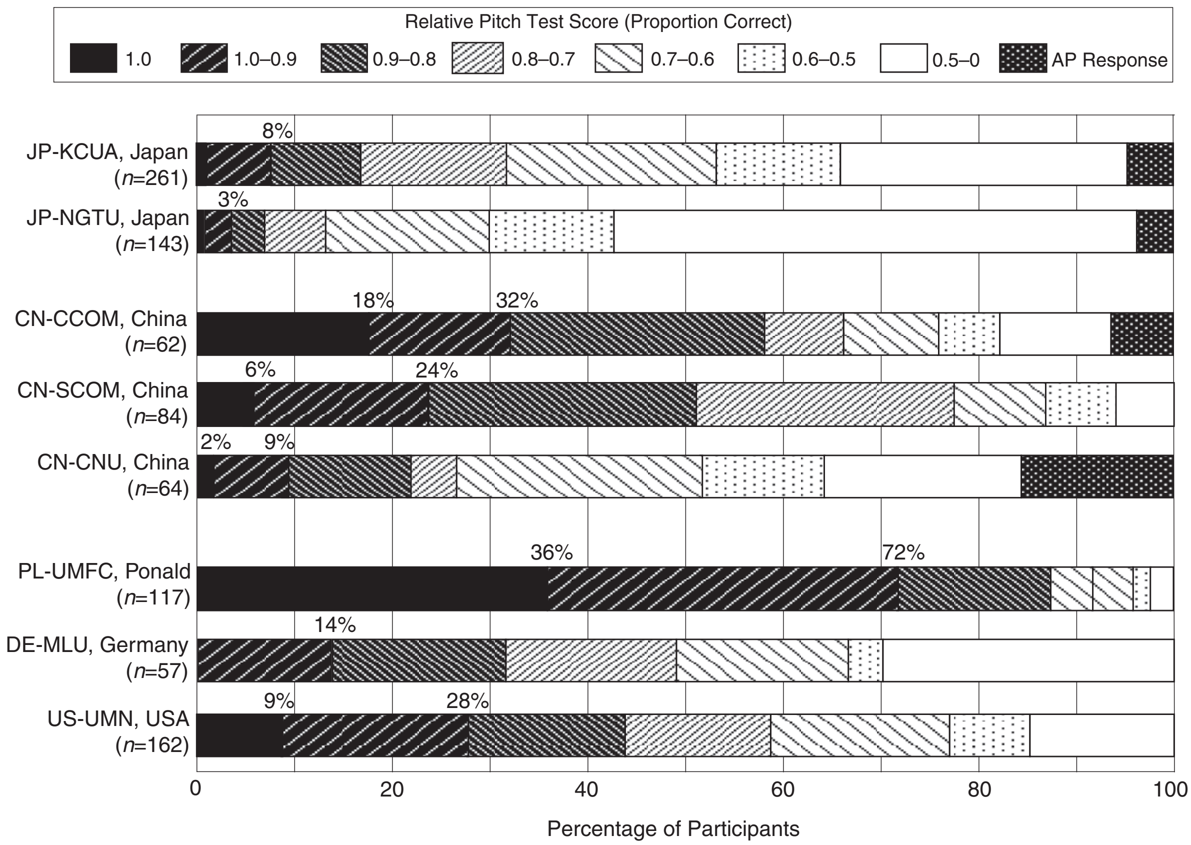



**FIGURE 4.** Percentages of the participants classified by their RP scores for each group. Participants majoring in Chinese traditional instruments at CN-SCOM and CN-CNU were excluded. The right-most area designated as AP Responses represents percentage of the participants who answered two AP names instead of the RP name required. Numbers above each column indicate percentages of participants who achieved a perfect score (1.0) and an accurate score (0.9 or above).

**TABLE 3.** *Median and Lower/Upper Quartile Values of the RP Scores (Proportion Correct) with the Summary of the Multiple Comparisons Among Groups*

| | Median | | | Homogeneous Subsets | | | |
|---|---|---|---|---|---|---|---|
| | $Q_1$ | $Q_2$ | $Q_3$ | | | | |
| JP-KCUA (Japan) | 0.485 | 0.606 | 0.758 | | | c | |
| JP-NGTU (Japan) | 0.364 | 0.485 | 0.606 | | | | d |
| CN-CCOM (China) | 0.697 | 0.818 | 0.932 | | b | | |
| CN-SCOM (China) | 0.727 | 0.818 | 0.879 | | b | | |
| CN-CNU (China) | 0.523 | 0.606 | 0.803 | | | c | |
| PL-UMFC (Poland) | 0.879 | 0.939 | 1.000 | a | | | |
| DE-MLU (Germany) | 0.485 | 0.697 | 0.818 | | | c | |
| US-UMN (USA) | 0.606 | 0.758 | 0.909 | | b | | |

Note: Participants majoring Chinese traditional music and participants who made AP responses were excluded from the analysis.

was markedly poorer. Most noteworthy is that the JP-KCUA students majoring in singing were poorest in RP identification; not a single participant in this group achieved the accurate level criterion and nearly half of them did not meet the minimum criterion (an RP score of 0.5). This result is in marked contrast to the subgroup majoring in singing at CN-SCOM, in which participants achieved fairly high performance in RP (45% of them at the accurate level), the highest score for CN-SCOM participants. The subgroups of the Chinese participants (CN-CCOM and CN-SCOM) generally performed better than the Japanese JP-KCUA group; interestingly, even the students of CN-SCOM who majored in Chinese traditional musical instruments (e.g., erhu, guzheng, pipa, and yangqin) performed fairly well in the RP task (37% of them at the accurate level). Compared with the Japanese and Chinese group, the subgroups of PL-UMFC (Poland) showed generally high performance. In PL-UMFC, a majority of the participants (54–95%) of all subgroups reached the accurate RP level; among others, almost all of the students who were Sound Engineering majors exceeded the accurate RP criterion.

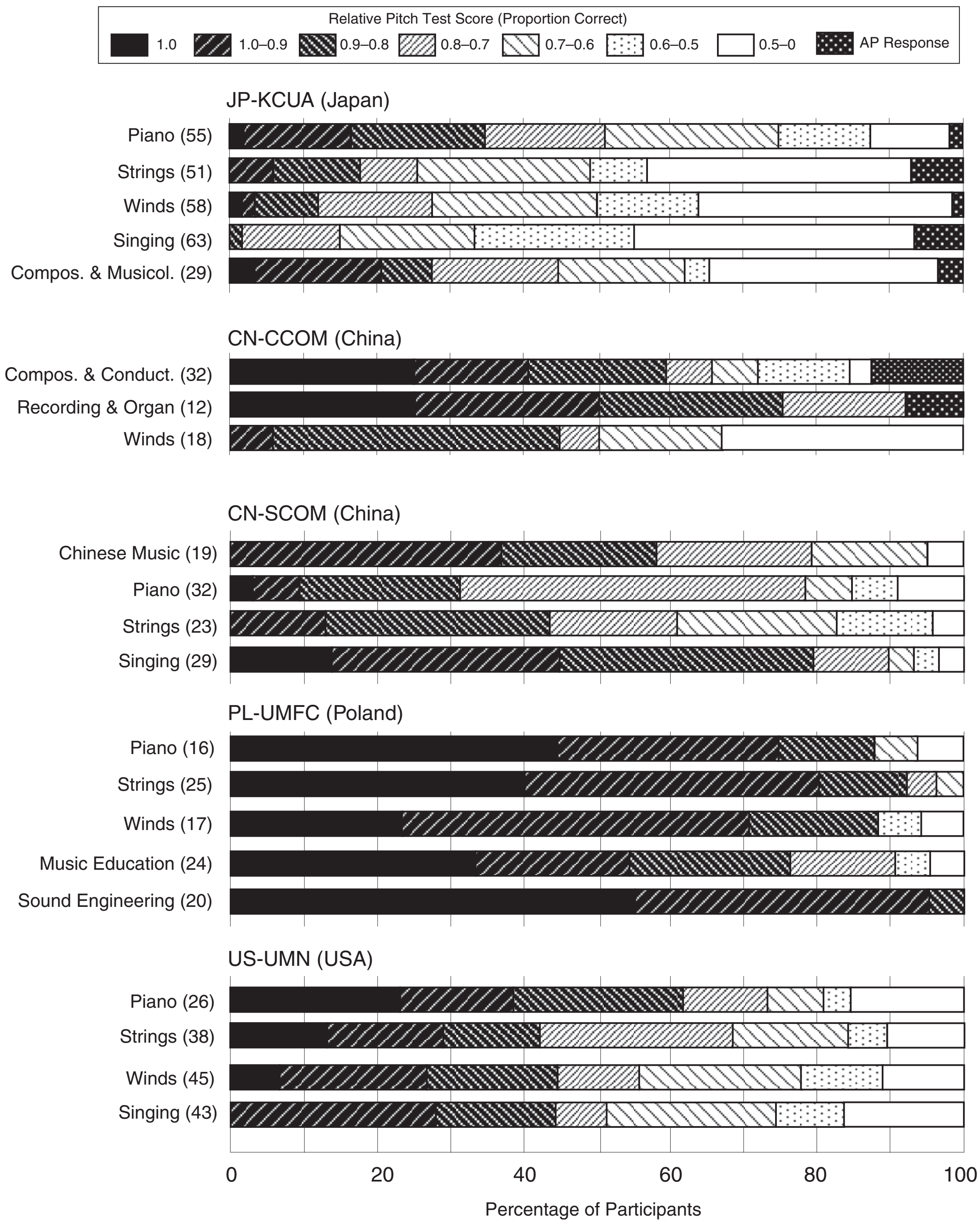



**FIGURE 5.** Percentages of the participants classified according to their RP scores for different specialties in each group.

Performance in the RP test should be dependent on the amount of aural-skills training the participants had received. The participants were continuing their aural-skills study, and so the performance of the RP test might improve if measured at a later time in their education. Figure 6 represents the average RP score by the semester in which each student was when taking the test; three groups (JP-KCUA, JP-NGTU, and US-UMN) are included, for which comparably sized semester subgroups could be formed. JP-KCUA and JP-NGTU

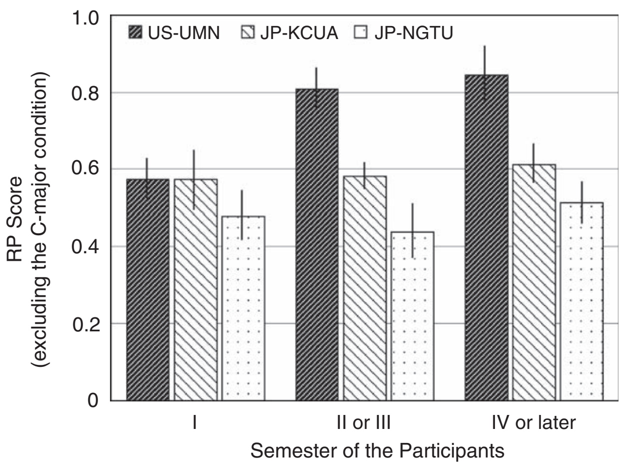



**FIGURE 6.** Average RP score of the participants in different semesters for US-UMN, JP-KCUA, and JP-NGTU. Error bars show 95% confidence intervals.

included subgroups of the first, third, and the fourth or later semesters, while US-UMN included those of the first, second and fourth semesters.

The US-UMN students showed a marked improvement of the RP score between the first and the second semester, whereas the JP-KCUA and JP-NGTU students showed no discernible improvements among semesters. This observation was borne out in a two-way analysis of variance (ANOVA) with institution and semester as between-subjects factors, which yielded a significant interaction between the factors, $F(4, 538) = 7.34$, $p < .0001$, $\eta_p^2 = .05$, as well as significant main effects of institution, $F(2, 538) = 53.14$, $p < .0001$, $\eta_p^2 = .17$, and semester, $F(2, 538) = 9.62$, $p < .0001$, $\eta_p^2 = .04$. Separate follow-up one-way ANOVAs revealed that an effect of semester was significant only for US-UMN, $F(2, 159) = 29.33$, $p < .0001$, $\eta_p^2 = .27$, in which the first semester subgroup was substantially lower in the RP score than the second and the fourth semester subgroups (0.58 vs. 0.81 and 0.85, respectively; $p < .05$ with Bonferroni correction for multiple comparisons) and the difference between the latter two subgroups was not significant. By contrast, there were no significant differences among all the semester subgroups in JP-KCUA and JP-NGTU ($\eta_p^2 = .004$ for JP-KCUA and $\eta_p^2 = .02$ for JP-NGTU). The main effect of institution was significant, $F(2, 538) = 53.135$, $p < .0001$, $\eta_p^2 = .17$, but significant difference among the institution subgroups was observed only in the second/third and the fourth or later semester categories, $F(2, 242) = 38.57$, $p < .0001$, $\eta_p^2 = .24$, and $F(2, 158) = 27.41$, $p < .0001$, $\eta_p^2 = .26$, respectively, in which the RP score of US-UMN was significantly higher than that of JP-KCUA and JP-NGTU ($p < .05$), but in the first semester US-UMN was similar to JP-KCUA and JP-NGTU, $F(2, 138) = 2.47$, $p > .05$, $\eta_p^2 = .04$.

### RELATIONSHIP BETWEEN AP AND RP PERFORMANCE

The analyses described so far clearly demonstrate an opposite pattern of AP and RP performance among participant groups. This contrasting relationship among groups is easily recognized in a scatterplot (Figure 7), in which individual groups are plotted in terms of their average AP scores on the horizontal axis and their average RP scores on the vertical axis. Individual groups are scattered representing, on the whole, an inverse relationship between the average scores of AP and RP. This scattered distribution of groups was due to the large differences in the amount and types of music training given to the participants. In particular, the conservatory-level groups (JP-KCUA, CN-CCOM, CN-SCOM, and PL-UMFC), represented by filled circles, lie on a line and represent an inverse relationship: JP-KCUA with a high AP score and a low RP score, PL-UMFC with a low AP score and a high RP score, and CN-CCOM and CN-SCOM with intermediate AP and RP scores. Other groups (represented by unfilled circles) scored lower both on the AP and the RP tasks than the conservatory-level groups, and also fall into a pattern that suggests an inverse (though less clear) relationship. Furthermore,

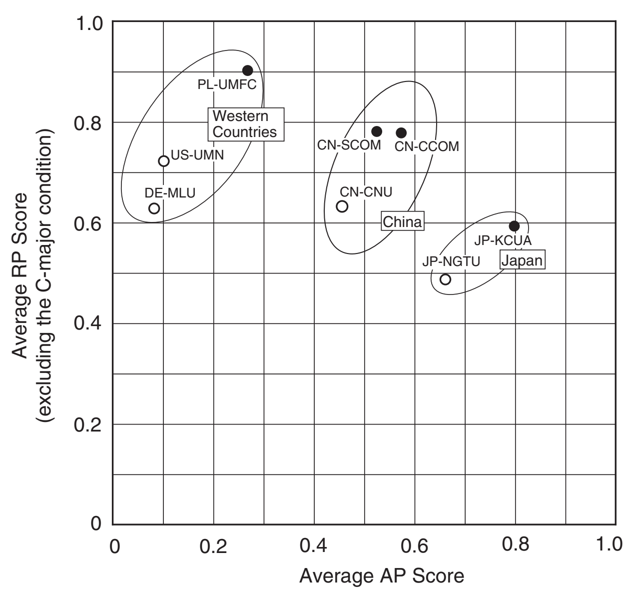



**FIGURE 7.** Scatterplot showing as a whole the relationship between AP and RP scores averaged across participants for each group. Filled symbols represent the conservatory-level music universities and unfilled symbols show other non-conservatory groups. The groups are combined into three larger groups (Japanese, Chinese, and European/American) enclosed in ovals. Participants majoring in Chinese traditional instruments at CN-SCOM and CN-CNU were excluded.

when the subgroups are combined into three larger groups (Japan, China, and Western countries), we can see the difference among these socio-cultural groups: the Japanese groups with excellent AP and poor RP, the Western groups with poor AP and excellent RP, and the Chinese groups with intermediate AP and RP.

When we assess the relationship between the AP and RP scores of the individual participants for each group separately, a different picture emerges. In Figure A1, individual participants are plotted according to their AP and RP scores for each selected group (JP-KCUA, CN-SCOM, PL-UMFC, CN-CNU, and JP-NGTU). Correlational analyses were performed using the AP and RP scores of the individual participants for each group, excluding those who responded errantly using AP responses. The results showed relatively small or non-significant correlation coefficients (positive or negative) for JP-KCUA: $r(249) = .30$, $p < .001$, CN-SCOM: $r(103) = -.25$, $p = .01$, CN-CCOM: $r(58) = .11$, $p = .41$, PL-UMFC: $r(117) = .001$, $p = .99$, CN-CNU: $r(84) = .049$, $p = .66$, and JP-NGTU: $r(136) = .08$, $p = .36$.

The correlations, positive and negative, might come from an incidental mixture of participants with particular specialties whose AP or RP scores were relatively higher or lower. Looking into the scatterplot representing JP-KCUA (Panel A in Figure A1), we find that, while a large number of participants are widely scattered, participants with perfect AP account for about a third of the JP-KCUA participants, who also achieved relatively high scores in RP, resulting in a relatively weak but significant positive correlation. CN-SCOM participants (Panel B) specializing in singing and Chinese traditional instruments whose AP scores were low, performed relatively well on the RP test, while those with high AP scores, many of whom were composition and piano majors, did not perform as well on the RP test, resulting in a weak negative correlation.

Finally, we subdivided participants of each group into two to four subgroups of statistically comparable sample size using their AP scores and further analyzed RP performance in different key conditions across those subgroups. The third column of Table 4 shows the average

**TABLE 4.** *Mean RP Scores and Mean Proportions Correct in All Key Conditions for Subgroups of Different AP Accuracy*

| AP Subgroups (AP Score Range) | *n* | RP Score | Key | | | |
|---|---|---|---|---|---|---|
| | | | C | E+ | F♯ | A♭ |
| JP-KCUA (Japan) | | | | | | |
| AP1 (1.0) | 86 | 0.713* | 0.876* | 0.689 | 0.718 | 0.732 |
| AP2 (.9–1.0) | 55 | 0.595 | 0.830* | 0.592 | 0.592 | 0.602 |
| AP3 (.5–.9) | 65 | 0.495 | 0.703* | 0.522 | 0.509 | 0.453 |
| AP4 (<.5) | 43 | 0.502 | 0.649* | 0.503 | 0.497 | 0.505 |
| JP-NGTU (Japan) | | | | | | |
| AP1 (1.0) | 23 | 0.57 | 0.822* | 0.593 | 0.549 | 0.569 |
| AP2 (.9–1.0) | 25 | 0.482 | 0.778* | 0.487 | 0.48 | 0.48 |
| AP3 (.5–.9) | 46 | 0.483 | 0.741* | 0.506 | 0.48 | 0.462 |
| AP4 (<.5) | 42 | 0.465 | 0.578* | 0.474 | 0.487 | 0.435 |
| CN-CCOM (China) | | | | | | |
| AP1 (.9–1.0) | 15 | 0.887 | 0.988* | 0.806* | 0.933 | 0.921 |
| AP2 (.5–.9) | 18 | 0.722 | 0.929* | 0.707 | 0.737 | 0.722 |
| AP3 (<.5) | 25 | 0.759 | 0.775 | 0.745 | 0.785 | 0.745 |
| CN-SCOM (China) | | | | | | |
| AP1 (.9–1.0) | 23 | 0.794* | 0.937* | 0.763 | 0.842 | 0.779 |
| AP2 (.5–.9) | 20 | 0.659 | 0.795* | 0.695 | 0.686 | 0.595 |
| AP3 (<.5) | 60 | 0.830* | 0.794 | 0.847 | 0.836 | 0.806 |
| CN-CNU (China) | | | | | | |
| AP1 (.8–.9) | 11 | 0.766 | 0.893* | 0.802 | 0.76 | 0.736 |
| AP2 (.5–.8) | 22 | 0.612 | 0.810* | 0.624 | 0.603 | 0.607 |
| AP3 (<.5) | 51 | 0.665 | 0.642 | 0.665 | 0.713 | 0.617 |
| PL-UMFC (Poland) | | | | | | |
| AP1 (.8–1.0) | 16 | 0.892 | 0.915 | 0.886 | 0.898 | 0.892 |
| AP2 (<.6) | 101 | 0.904 | 0.889 | 0.91 | 0.897 | 0.905 |

Note. The RP score is a combination of the E+, F♯, and A♭ conditions. The AP score range of each subgroup (in parentheses) is lower inclusive. An asterisk in the RP Score column indicates the performance of the subgroup was significantly higher than the other subgroup(s) within that group. An asterisk in the C column indicates the performance in the C major condition was significantly higher than the combined other key conditions in each AP subgroup. An asterisk in the E+ column indicates the performance in the E+ major condition was significantly lower than the combined F♯ and A♭ major conditions in each AP subgroup.

RP scores of the subgroups of different AP accuracy in each group separately. In accordance with the correlational analyses, ANOVAs revealed significant effects of AP subgroups only for JP-KCUA, $F(3, 245) = 16.95$, $p < .001$, $\eta_p^2 = .17$, and CN-SCOM, $F(2, 100) = 10.44$, $p < .001$, $\eta_p^2 = .17$. Post hoc analyses using Tukey HSD showed that, in JP-KCUA, the RP performance was significantly better for AP1 (perfect AP) than other subgroups and for AP2 (accurate AP) than AP3 (inaccurate AP), and in CN-SCOM AP3 (non-AP) and AP1 (accurate AP) were better than AP2 (intermediate AP; $p < .05$). There was no statistically significant difference among subgroups at other music schools.

The RP test included four different key conditions in which the tonic was C, F♯, A♭, or E+ (a quarter-tone higher E than the standard pitch). The proportions correct in these four key conditions across AP subgroups for each group are shown in the right four columns in Table 4. AP is supposed to function as a beneficial tool for its possessor in the C major condition of the RP test, which was a reason why we excluded this condition in calculating the RP scores. As expected, the performance in the C major condition was better than other key conditions in many subgroups. A mixed-design ANOVA using a within-subjects factor of Key and a between-subjects factor of AP Subgroup conducted separately for each group indicated that the main effect of Key and the Key by AP Subgroup interaction were significant for all groups except for PL-UMFC. Planned comparisons by *t*-tests showed that the performance in the C major condition was significantly higher than performance in the other key conditions combined in all the subgroups of JP-KCUA and JP-NGTU, AP1 and AP2 of CN-CCOM, CN-SCOM and CN-CNU ($p < .05$, $d = 0.51$–$1.61$). In the E+ condition, in which AP listening might cause confusion because all tones were equally mistuned, the assumed interfering effect of mistuning was limited; only AP1 of CN-CCOM showed a significant decline in performance when compared to the combined F♯ and A♭ conditions, $t(14) = -2.78$, $p = .02$, $d = .64$.

## Discussion

We examined the AP and RP abilities of a large number of music students attending higher education institutions for professional musicians and music teachers in different countries from East Asia and the West. The results demonstrated a striking contrast of the AP and RP performance between the East Asian and the Euro/American countries. A large proportion of participants had accurate AP in Japan and, to a lesser extent, in China, while in the Western countries, only a small minority, if any, had AP. The results of the RP test showed an opposite pattern; as a whole, the Japanese participants were very poor in RP performance, while the Western participants performed well and the Chinese participants were intermediate. Thus, there was an inverse relationship between the average AP and RP scores among the institutions to which the participants belonged. However, when plotting the AP and RP scores of the individual participants within each group separately, almost no correlation, or only a weak positive or negative correlation, was observed.

### THE DIFFERENCE IN AP PREVALENCE AND ITS ORIGIN

Results of the present AP test were largely consistent with the previous study regarding the prevalence of AP. Specifically, the distribution of AP accuracy in JP-NGTU (Japan) and PL-UMFC (Poland) closely replicated our previous findings (Miyazaki et al., 2012). Furthermore, the percentage of accurate level AP participants in the composition and conducting group of CN-CCOM (about 60% with at least 0.85 correct) was close to that of CN-CCOM (about 55%) previously reported by Deutsch et al. (2006), while the overall AP score of CN-SCOM participants (0.53) was slightly lower than that of CN-SCOM (0.68) calculated from a figure in Deutsch et al. (2013). However, the results of these comparisons may not be taken at face value because of the difference in the stimulus condition of the AP test used and in the constitution of the participants of different music specialties between the present study and Deutsch et al.'s studies. On the other hand, in the present study, the percentage of accurate AP participants at PL-UMFC was much lower than that of the Japanese and Chinese groups. This is consistent with the previous notion that there are fewer music students who possess AP in the West.

The present study provided a new finding: the percentage of AP possessors was significantly higher in Japan than in China, both for conservatory-level training—JP-KCUA (Japan) vs. CN-CCOM and CN-SCOM (China)—and in music education programs—JP-NGTU (Japan) vs. CN-CNU (China). More specifically, the average AP score was higher for all of the subgroups of different specialties in JP-KCUA than the corresponding subgroups in CN-CCOM or CN-SCOM, in which there were only a small percentage of AP participants majoring in wind instruments and singing. These results suggest that AP prevalence is higher in Japan than in China.

Why is AP so prevalent among music students in Japan and, to a lesser extent, in China in comparison to Europe and America? The marked difference in the prevalence of AP may be explained partly by a difference in genetic contributions. Some supporting evidence for a genetic basis of AP has been reported in familial aggregation studies (Baharloo, Johnston, Service, Gitschier, & Freimer, 1998; Baharloo, Service, Risch, Gitschier, & Freimer, 2000; Gregersen, Kowalsky, Kohn, & Marvin, 1999, 2000) and twin studies (Theusch & Gitschier, 2011). Furthermore, a genome-wide study (Theusch, Basu, & Gitschier, 2009) provided preliminary evidence revealing specific genes that contribute to AP and suggested that AP is a complex and genetically heterogeneous trait. Unfortunately, at present, the evidence for a genetic component for AP is limited and not sufficiently convincing. Indeed, there is no doubt that all human traits have a genetic foundation to a greater or lesser extent, and, therefore, the possibility that a genetic basis for AP may be prevalent in some East Asian populations cannot be ruled out. However, to determine this definitively, the genetic mechanisms that allow AP to evolve must be elucidated. Moreover, we should note that any phenotypes would be determined by complex interactions between genetic predispositions and a number of environmental factors.

As one of the environmental factors, influences of language experience early in life may contribute to the later development of AP. Deutsch and her colleagues proposed a tone-language hypothesis to explain the high prevalence of AP in China (Deutsch, Dooley, Henthorn, & Head, 2009; Deutsch, Henthorn, & Dolson, 2004; Deutsch et al., 2006). They argued that in tonal languages such as Mandarin and Cantonese, lexical tone defined both by pitch height and by pitch contour determine the meaning of words, and early exposure to these languages facilitates the development of AP. However, we should be cautious about drawing a conclusion from the data presently in hand. The primary data provided in support of the tone-language hypothesis are correlational and thus do not rule out the possible contributions of other factors correlating with the tone-language proficiency (but see Deutsch et al., 2009). Moreover, the lexical tones are considered to be defined not by absolute pitch but by relative pitch in a broader sense; apart from pitch contour, which is explicitly based on relative pitch, another component defining the lexical tones, is not strictly defined by absolute pitch but rather loosely defined by pitch register relative to the speaker's vocal range (Peng, 2006; Peng & Wang, 2005). Furthermore, Japanese is not a tonal language but a pitch-accent language that is located at an intermediate position between a typical stress language (such as English) and a typical tonal language on the tonal density continuum (Hyman, 2009). If tone-language speakers have an advantage in developing AP, one would predict that the prevalence of AP should have been higher among the Chinese group than the Japanese group. Actually, the results of the present study were opposite to this prediction from the tone-language hypothesis of AP. We would need other explanations for the Japanese advantage in AP.

The higher prevalence of AP in Japan is more likely due to early music training widespread in Japan that facilitates the development of AP. Our Japanese groups initiated music training about 1.5 to 2.5 years earlier than other groups. This difference is very significant, because early music training from 3 to 6 years old is most effective for AP acquisition (Miyazaki et al., 2012; Sakakibara, 2014; see Deutsch, 2013; Takeuchi & Hulse, 1993, for review). Our results showed that AP performance and the percentage of accurate AP participants decreased as a function of the age of training onset, supporting this view regarding AP acquisition. It should be noted, however, that, in JP-KCUA, the early training effect was indeed evident but the percentage of accurate AP possessors still remained relatively high in the participants whose age of training onset was 7 and 8 years or later (50% and 30%, respectively). This suggests the possibility that the period in which AP is effectively learned may not so strictly determined as the critical period is presumed to be but, rather, the age window for AP development may be somewhat flexible.

Although the age of onset for music training is the influence most frequently mentioned in the AP literature, early music training does not necessarily ensure AP development. More important, it is not the early training per se but the content of this training. Specifically, the fixed-*do* singing practice purposefully given for AP acquisition is supposed to be effective in the formation of an association between individual pitches and their pitch labels (Wilson, Lusher, Martin, Rayer, & McLachlan, 2012). In fact, there are a large number of music schools for children throughout Japan (e.g., the Yamaha Music Schools and multitudes of smaller individual music schools) that emphasize the importance of early education for musical development and promote AP training. In the Yamaha Music Schools, children generally start by singing songs using fixed-*do* syllables, listening to or playing piano (see Miyazaki & Ogawa, 2006, for more detail), and some of these children are presumed to acquire AP as a result of this type of music training. There are even music schools focusing on AP training, in which children take lessons using a method specially designed for learning AP (see Sakakibara, 2014, for more detail). A longitudinal study conducted in an AP school reported that all children who had begun lessons at 3 years of age and had strictly followed instructions in the "chord identification method" successfully developed AP (Sakakibara, 2014). Behind the flourishing of early AP training in Japan, there seems to be a sociocultural belief that values AP as a sign of musical prowess.

Summarizing the discussion about AP so far, there may be multiple factors interacting with each other that contribute to AP development: a genetic predisposition probably providing the basis for the expression of AP, early exposure to a tonal language partly contributing to later AP development (specifically as in case of Chinese musicians), early learning with specific training methods leading to AP acquisition, and the socio-cultural belief favoring AP (as in Japan).

### RP PERFORMANCE AND ITS RELATION TO AP PERFORMANCE

The most remarkable finding of the present study is an opposite pattern of the distribution of the AP and RP accuracy among the groups of music students in East Asia and the West: the Japanese groups with the highest AP scores had the lowest RP scores, whereas the Western groups with the lowest AP scores had the highest RP scores. This contrast indicates that excellent AP may not have a positive impact for musicians in their cognitive processing of RP, which is opposite to the prevailing view favoring AP as a valued ability for musicians. However, despite an inverse relationship between the average AP and RP scores of the individual groups (Figure 7), the correlation of AP and RP scores of individual participants in each group yielded mixed results, with weak or negligible correlation, positive or negative (Figure A1).

It should be noted that it is difficult to measure RP directly. A musical interval is defined by two notes and, therefore, participants who have accurate AP could have an advantage, allowing somewhat higher performance on the RP test in comparison to those relying solely on RP ability without accurate AP; they would be able to identify the two pitches presented without effort using AP and promptly determine the interval between them. However, the strategy of relying on AP, no matter how beneficial it may be in isolated contexts, violates the intent of the RP test to measure an ability to perceive the tonal function of each scale-degree tone in a musical context, and, therefore, causes a decrease in test validity.

To suppress the AP strategy being used in the RP test, we provided the participants with instructions stressing that they should hear the chord-tone pair sequences in a specific key context and identify the scale-degree of the target tone. Nevertheless, it is likely that the AP strategy was not fully suppressed and resulted in higher RP scores for participants with accurate AP. It could have provided an automatic and familiar way for some participants with accurate AP to perform the RP task, capitalizing on their AP ability. In a brain imaging study (Wilson, Lusher, Wan, Dudgeon, & Reutens, 2009), activities corresponding to automatic AP processing were observed when AP musicians were engaged in RP processing. In the present study, in fact, some Japanese and Chinese participants not only answered the RP names required but also wrote down the AP names of the tone pairs presented as annotations in the margins of their response sheets. Such markings provide evidence that they employed AP strategies in the RP test despite the instructions given. There could be other, more implicit, cases in which participants addressed the RP test using AP strategies without writing down AP names, perhaps mentally counting the number of scale steps between the two notes on an imagined keyboard or in mental notation. These participant responses would have been determined as if musical intervals are to be counted from the two pitches identified using AP rather than being aware of the musical pitch relations and tonal functions of the tones of the pitch pairs. We do not know exactly how many participants performed the RP test in this manner, since this process occurs invisibly in their minds. We asked the participants after the RP test session which strategy (AP or RP) they had used in the RP test. Responses to this question indicated that participants who reported they had used exclusively or partly the AP strategy accounted for 83% of the participants in JP-KCUA, 61% in CN-CCOM, 44% in CN-SCOM, 14% in PL-UMFC, 67% in JP-NGTU, 46% in CN-CNU, 14% in US-UMN, and 2% in DE-MLU. These self-reports about the strategy the participants used should be considered with reservation, because it is not certain whether the participants really understood the meaning of AP and RP strategies in the RP test. Nevertheless, it is possible that a considerable proportion of the Japanese and the Chinese participants performed the RP test using AP strategies.

This speculation is supported by the key effect: performance in the C major condition, in which the AP strategy was most effective, was significantly higher than performance in the other key conditions in the Japanese and Chinese groups (Table 4). We excluded the C major condition in calculating the RP scores, but the AP strategy might have influenced responses also in the other key conditions. The participants who employed AP strategies in the RP test despite the instructions probably did so because it was an optimal means for them to address the challenge of the RP test, which possibly resulted in somewhat higher RP scores than might have occurred had they limited themselves only to using RP ability as instructed.

We hypothesized that the AP strategy might not work well in the E+ condition, in which all tones were shifted by a quarter-tone from the pitch standard and would have presumably sounded out-of-tune to participants with AP. This mistuning was supposed to cause interference when identifying the notes in this condition, decreasing the RP performance of AP participants. However, the results showed that the assumed interfering effect of mistuning was relatively limited. Participants who used the AP strategy might have managed to cope with a quarter-tone shift by mentally adjusting their AP somehow. This speculation dovetails with an observation that some participants who answered AP names (AP responses) instead of RP names consistently wrote down the tonic of the E+ condition as E, F, or E♯.

From the discussion so far, one might conclude that at least some participants with AP could achieve somewhat higher RP performance with the aid of AP. Although this method of responding (i.e., relying on AP) is almost like cheating in light of the purpose of the RP test, the effect of this AP strategy for accurate AP possessors might have resulted in a positive correlation between the RP and the AP scores.

Furthermore, the level of musical experience of the participants may have influenced both the AP and the RP scores as a confounding variable, possibly producing a positive correlation between the RP and the AP performance (the third variable problem). In the present study, participants were students majoring in different musical specializations (piano, other instruments, singing, composition, and musicology), and so their musical experiences were different both in degree and quality. Generally, music students with AP began their music training earlier (Deutsch et al., 2013; Miyazaki et al., 2012; Sergeant, 1969; Wilson et al., 2012) and have had longer and more extensive musical experiences than those without AP. Moreover, AP possessors might have been more encouraged to go into a career in music, since they were likely to have been regarded as more talented as a result of their AP ability by their parents and music teachers, and they themselves might have had higher motivation to work hard in music lessons. Therefore, dividing participants into an AP group and a non-AP group inevitably produces a difference in the degree of musical experiences between them, and the AP group would perform better than the non-AP group in some musical tasks, unless the factor of musical experience was adequately controlled. This speculation is consistent with the results obtained from JP-KCUA in the present study, in which the perfect AP group showed significantly higher RP scores than other groups; this conclusion is also supported by the results from CN-CCOM, JP-NGTU, and CN-CNU, though not statistically significant (see Table 4, RP Score). These results coincide with those of Dooley and Deutsch (2011).

In short, the present study contained the confounding factors (the use of AP strategies and the differences in musical experience among participants). They may possibly result in a positive correlation between the AP and RP scores, which may appear to suggest a musical advantage of AP. However, our results revealed only a weak positive correlation (JP-KCUA), almost no correlation (CN-CCOM, PL-UMFC, JP-NGTU, and CN-CNU), and even a weak negative correlation (CN-SCOM). These results are generally consistent with our weak hypothesis that there is no significant correlation between AP and RP, implying that AP has less relevance to musical pitch processing than has been presumed. Furthermore, the results suggest the possibility that we could have observed negative correlations between the AP and RP scores if the confounding factors had been adequately controlled. The truth may be that AP musicians are inferior to non-AP musicians with equivalent musical experience in performing the RP task. This is consistent with the results of a previous experiment (Miyazaki, 1995), in which the exact same RP test was conducted in a laboratory setting with participants' musical experience controlled, demonstrating that the AP group performed worse in the RP test than the non-AP group. The weak or negligible correlations between the AP and RP performances observed in the present study may be a result of a mixture of the interfering effect of AP on RP processing, producing a negative correlation, and the opposite effects of the supposed confounding factors, producing a positive correlation. Along this line of thinking, our results are consistent with our strong hypothesis that AP is inversely related to RP, implying that AP may interfere with RP processing in a musical context.

One might argue that the interference of AP on RP processing was a result of a Stroop-like conflict engendered in the AP participants who had been trained in the fixed-*do* pitch-naming system by a mismatch between a fixed-*do* AP name and a movable-*do* RP name. Dooley and Deutsch (2011) proposed this argument when critiquing the results of the previous experiment by Miyazaki (1995). Deutsch (2013) commented that the task employed by Miyazaki (1995) was an unusual one (i.e., the unusual requirement to use sol-fa names to label musical intervals produced a Stroop-like situation), so that AP possessors would be expected to experience confusion in performing this task. However, the Stroop-like conflict, if it really occurred, would in itself qualify as evidence that the AP participants had not developed sufficient RP skills or could not control their AP to ignore the conflict. We consider such conflict unlikely to have occurred in both the previous and the present experiments, since, in both studies, participants were allowed to use either movable-*do* labels, musical interval labels, or any other familiar RP labels when responding in the RP procedure. If the participants had AP and had been trained in the fixed-*do* tradition, they could use the interval labels instead of the movable-*do* labels. It is quite unlikely that, in this situation, those participants would have chosen to use the movable-*do* labels that might have produced a conflict with a familiar fixed-*do* label. This was confirmed by the labels the participants actually used in the RP test. Only a small percentage or none answered using sol-fa names: 12% in JP-KCUA, 19% in JP-NGTU, and 0% in the other groups. Interestingly, the percentage was inversely related to the AP score; in JP-KCUA, it was 4% for the accurate AP group, 8% for the inaccurate AP group, and 42% for the non-AP group; in JP-NGTU, it was 4%, 19%, and 35%, respectively. Most of the Japanese participants and almost all of the Chinese and the Western participants answered using interval names.

It is often pointed out that musicians with AP may encounter problems caused by interference between AP processing and musical pitch processing based on RP, such as when transposing a well-remembered piece of music into another key, playing music on a keyboard instrument tuned to a historical pitch standard, or adapting to an incidental pitch drift. However, AP would lead to these problems only if musicians have AP at an automatic level but have not developed RP abilities sufficiently. Many musicians with AP are presumed to have made particular effort to attain proficiency in RP processing and an ability to appropriately control AP, so that they could overcome these challenges and take advantage of AP as a useful tool.

Rather, the real problem of AP would be the possibility that AP may interfere with the proper development of RP in a broader context, which needs to be considered from the developmental and socio-cultural perspectives. From a developmental perspective, AP and RP are entirely different modes of pitch processing that follow dissimilar developmental paths; AP is a simpler and more primitive pitch processing mode that develops in early childhood. According to the early learning view of AP, it is most efficiently acquired if some specific training is given to young children from 3 to 6 years of age (Miyazaki & Ogawa, 2006; Sakakibara, 2014; see Deutsch, 2013, Takeuchi & Hulse, 1993, for review). In contrast, RP entails more advanced pitch processing that requires explicit and/or implicit knowledge of the musical pitch context (tonal schema or tonality) and, hence, develops later than AP. This may be in line with the notion of the more general developmental shift from absolute to relational processing (Saffran & Griepentrog, 2001; Sergeant & Roche, 1973; Takeuchi & Hulse, 1993).

The pedagogical problem that arises in relation to AP lies in the process of musical development in the socio-cultural context of music education. Once children have acquired AP in the early years of their lives, they might become dependent on the acquired AP ability in their musical activities, because AP could be a useful tool for some musical activities. For example, AP is indeed very useful for making it through aural-skills tasks such as music dictation or sight-singing. However, it should be noted that accomplishing these tasks using AP is a shortcut and might undermine the aural-skills training that is provided primarily to cultivate RP ability. It would be difficult for small children to understand the real purpose of aural-skills training and the importance of RP, unless adequate explanation and modeling have been provided by informed teachers. If children with AP insist on relying on AP in aural-skills training, they might remain unaware of the value of relative-pitch training and aural-skills training would be inefficient for AP possessors. As a result, they may fail to develop critical RP skills sufficiently.

The problem is also on the side of music teachers. A teacher in an aural-skills class in which most students have AP might possibly be misled by the apparently high aural-skills performance accomplished by the use of AP, assuming it is proof that students have already acquired a high degree of aural-skills and erroneously concluding that they need no further training. As a consequence, at least some of the children or students with AP might proceed without sufficiently developed RP and without enough understanding about its musical importance. Additionally, in an AP-dominated class, the minority of non-AP students may not be provided with a deserved opportunity to develop their aural-skills, because the misguided teacher might evaluate them simply as incapable students. This speculation is consistent with the results shown in Figure 6, in which US-UMN (USA) participants show significant improvement in RP performance between the first and the second semesters. This improvement is interpreted to be results of effective aural-skills training the US-UMN students received during the first semester. In contrast, RP performance of all semester groups at JP-KCUA and JP-NGTU (both in Japan) remain low with no improvement between the semester groups, suggesting the possibility that aural-skills training given to the Japanese participants may be ineffective.

Behind the disregard for RP is a social belief that AP is a valuable musical ability, representing a higher form of musicianship. This is the same socio-cultural background that has been behind the flourishing of AP training, especially in Japan and China. As argued in the present paper, this notion is misleading but commonly accepted in some cultures and represents the status quo particularly in Japan. The zeal for this ability facilitates the continuing push for AP training, and concurrently may produce music students with poor RP. The attitude towards AP is different from culture to culture and this difference is presumed to influence the music-educational content in each country. The inverse relationship between AP and RP scores observed among groups of different cultures (Figure 7) is interpreted to reflect differences in the socio-cultural conditions.

To address the potential problems caused by AP in music education practice, children or music students who have acquired AP must be provided with adequate training of a remedial nature, specifically designed to cultivate RP ability. The fixed-*do*-based solfège practice strongly associated with AP is prevalent in Japan, but it does not work well for developing the RP ability. To overcome these problems, theories and practices of the movable-*do*-based Kodály method (e.g., Dobszay, 2009, for recent developments of the Kodály method) or the method using scale-degree numbers (e.g., Marvin, 2007) might be instructive. To address the importance of relative-pitch processing for AP students in their music education in the long run, it is necessary to dispel the myths and misconceptions about AP (Miyazaki, 2014).

## Author Note

Co-author Andrzej Rakowski passed away during the last stages of the review process for this article.

This research was supported by Grant-in-Aid for Scientific Research (C) (No. 21500256) from the Japan Society for the Promotion of Science to the first author. Part of this article was presented at the Conference of the European Society for the Cognitive Science of Music (ESCOM), August 2015, Manchester, UK, and at the International Congress of Psychology, July 2016, Yokohama, Japan. We thank the action editor and two anonymous reviewers for their helpful comments and suggestions on this work.

*Correspondence concerning this article should be addressed to* Ken'ichi Miyazaki, Faculty of Humanities, Niigata University, 8050 Ikarashi-2-no-cho, Nishi-ku, Niigata, 950-2181, Japan. E-mail: miyazaki@human.niigata-u.ac.jp

## References

Bachem, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America, 9*, 146–151.

Bachem, A. (1940). The genesis of absolute pitch. *Journal of the Acoustical Society of America, 11*, 434–439.

Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America, 27*, 1180–1185.

Baharloo, S., Johnston P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics, 62*, 224–231.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *American Journal of Human Genetics, 67*, 755–758

Chavarria-Soleya, G. (2016). Absolute pitch in Costa Rica: Distribution of pitch identification ability and implications for its genetic basis. *Journal of the Acoustical Society of America, 140*, 891–897.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science*, *11*, 200–204.

Deutsch, D. (2013). Absolute pitch. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 141–182). San Diego, CA: Elsevier.

Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American music conservatory: Association with tone language fluency. *Journal of the Acoustical Society of America, 125*, 2398–2403.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception, 21*, 339–356.

Deutsch, D., Henthorn, T., Marvin, E. W., & Xu, H.-S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America, 119*, 719–722.

Deutsch, D., Li, X., & Shen, J. (2013). Absolute pitch among students at the Shanghai Conservatory of Music: A large-scale direct-test study. *Journal of the Acoustical Society of America, 134,* 3453–3859.

Dobszay, L. (2009). *After Kodály: Reflections on music education.* Kecskemét: Kodály Institute of the Liszt Academy.

Dooley, K., & Deutsch, D. (2010). Absolute pitch correlates with high performance on musical dictation. *Journal of the Acoustical Society of America, 128,* 890–893.

Dooley, K., & Deutsch, D. (2011). Absolute pitch correlates with high performance on interval naming tasks. *Journal of the Acoustical Society of America, 130,* 4097–4104.

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (1999). Absolute pitch: Prevalence, ethnic variation, and estimation of the genetic component. *American Journal of Human Genetics, 65*, 911–913.

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (2000). Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment. *American Journal of Medical Genetics, 98*, 280–282.

Hyman, L. M. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Science, 31*, 213–238.

Karpinski, G. S. (2000). *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians*. New York: Oxford University Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Krumhansl, C. L. & Cuddy, L. L. (2010). A theory of tonal hierarchies in music. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception. Springer handbook of auditory research 36* (pp. 51–87). New York: Springer.

Leite, R. B. C., Mota-Rolim, S. A., & Queiroz, C. M. T. (2016). Music proficiency and quantification of absolute pitch: A large-scale study among Brazilian musicians. *Frontiers in Neuroscience, 10*, 447.

Lerdahl, F. (2001). *Tonal pitch space*. New York: Oxford University Press.

Lerdahl, F. (2015). Concepts and representations of musical hierarchies. *Music Perception, 33,* 83–95.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception, 24,* 329–366.

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Science, 9*, 26–33.

Marvin, E. W. (2007). Absolute pitch perception and the pedagogy of relative pitch. *Journal of Music Theory Pedagogy, 21*, 1–38.

Miyazaki, K. (1993). Absolute pitch as an inability: Identification of musical intervals in a tonal context. *Music Perception, 11,* 55–72.

Miyazaki, K. (1995). Perception of relative pitch with different references: Some absolute-pitch listeners can't tell musical interval names. *Perception and Psychophysics, 57,* 952–970.

Miyazaki, K. (2004). Recognition of transposed melodies by absolute-pitch possessors. *Japanese Psychological Research, 46*, 270–282.

Miyazaki, K. (2014). *Zettai onkan shinwa* [Myths of absolute pitch: Science reveals the true pictures]. Kyoto: Kagaku-Dojin.

Miyazaki, K., Makomaska, S., & Rakowski, A. (2012). Prevalence of absolute pitch: A comparison between Japanese and Polish music students. *Journal of the Acoustical Society of America, 132*, 3484–3493.

Miyazaki, K., & Ogawa, Y. (2006). Learning absolute pitch by children: A cross-sectional study. *Music Perception, 24*, 63–78.

Miyazaki, K., & Rakowski, A. (2002). Recognition of notated melodies by possessors and nonpossessors of absolute pitch. *Perception and Psychophysics, 64,* 1337-1345.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguist, 34*, 134–154.

Peng, G., & Wang, W. S.-Y. (2005). Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication, 45*, 49–62.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology, 37*, 74–85.

Sakakibara, A. (2014). A longitudinal study of the process of acquiring absolute pitch: A practical report of training with the 'chord identification method.' *Psychology of Music, 42*, 112–127.

Sergeant, D. (1969). Experimental investigation of absolute pitch. *Journal of Research in Music Education, 17*, 135–143.

Sergeant, D., & Roche, S. (1973). Perceptual shifts in the auditory information processing of young children. *Psychology of Music, 1*, 39–48.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin, 113*, 345–361.

Theusch, E., Basu, A., & Gitschier, J. (2009). Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity. *American Journal of Human Genetics, 85*, 112–119.

Theusch, E., & Gitschier, J. (2011). Absolute pitch twin study and segregation analysis. *Human Genetics, 14*, 173–178.

Ward, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 265-298). New York: Academic Press.

Wilson, S. J., Lusher, D., Wan, C. Y., Dudgeon, P., & Reutens, D. C. (2009). The neurocognitive components of pitch processing: Insights from absolute pitch. *Cerebral Cortex, 19*, 724–732.

Wilson, S. J., Lusher, D., Martin, C. L., Rayner, G., & McLachlan, N. (2012). Intersecting factors lead to absolute pitch acquisition that is maintained in a "fixed do" environment. *Music Perception, 29,* 285–296.

# Appendix

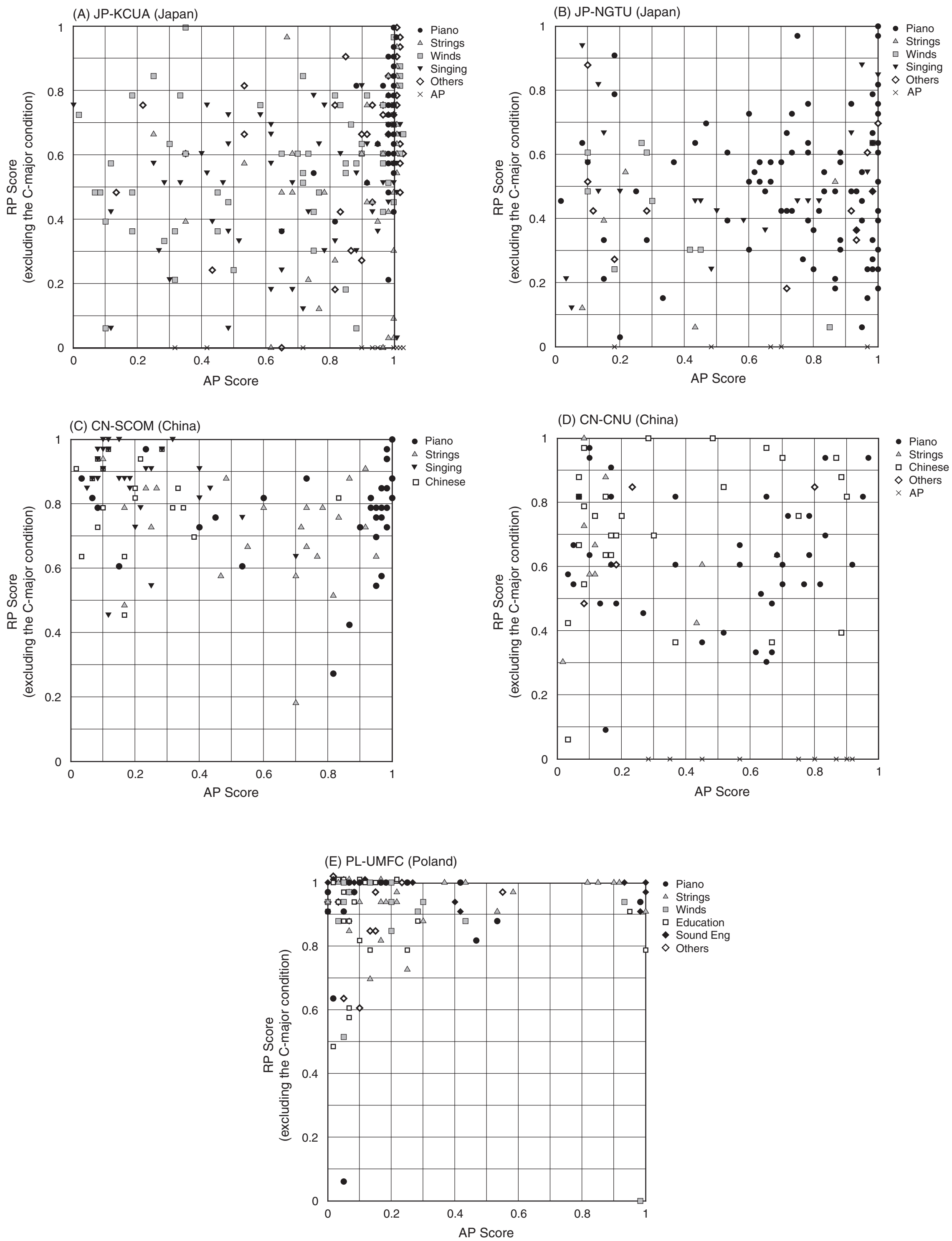



FIGURE A1. The relationship between AP and RP scores for individuals in selected groups. Different symbols in each panel represent individual participants in subgroups of different specialties. Some symbols of the perfect AP in JP-KCUA (A) and the perfect RP in PL-UMFC (E) are offset horizontally or vertically so that overlapped points of different subgroups are visible, but the overlap of data points in the same subgroup are not represented to avoid an overly complicated appearance.