Available online at www.sciencedirect.com

ScienceDirect

Hearing Research 219 (2006) 36–47

Hearing Research

www.elsevier.com/locate/heares

Research paper

Influence of musical and psychoacoustical training on pitch discrimination

Christophe Micheyl [a,*], Karine Delhommeau [b,c], Xavier Perrot [d], Andrew J. Oxenham [a,e]

[a] MIT Research Laboratory of Electronics, Massachusetts Institute of Technology, 50 Vassar Street, Bldg 36-758, Cambridge, MA 02139-4307, USA
[b] UMR CNRS 7593, Pavillon Clérambault, La Pitié-Salpêtrière, Paris, France
[c] Neuropsychology Department, Montreal Neurological Institute, Montreal, Canada
[d] UMR CNRS 5020, Université Claude Bernard Lyon 1, Lyon, France
[e] Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA

Received 15 July 2005; received in revised form 16 May 2006; accepted 18 May 2006
Available online 12 July 2006

Abstract

This study compared the influence of musical and psychoacoustical training on auditory pitch discrimination abilities. In a first experiment, pitch discrimination thresholds for pure and complex tones were measured in 30 classical musicians and 30 non-musicians, none of whom had prior psychoacoustical training. The non-musicians' mean thresholds were more than six times larger than those of the classical musicians initially, and still about four times larger after 2 h of training using an adaptive two-interval forced-choice procedure; this difference is two to three times larger than suggested by previous studies. The musicians' thresholds were close to those measured in earlier psychoacoustical studies using highly trained listeners, and showed little improvement with training; this suggests that classical musical training can lead to optimal or nearly optimal pitch discrimination performance. A second experiment was performed to determine how much additional training was required for the non-musicians to obtain thresholds as low as those of the classical musicians from experiment 1. Eight new non-musicians with no prior training practiced the frequency discrimination task for a total of 14 h. It took between 4 and 8 h of training for their thresholds to become as small as those measured in the classical musicians from experiment 1. These findings supplement and qualify earlier data in the literature regarding the respective influence of musical and psychoacoustical training on pitch discrimination performance.

Keywords: Musicians; Pitch; Frequency discrimination; Auditory training; Perceptual learning

1. Introduction

While a mystique surrounds the auditory abilities of musicians, relatively few studies have been devoted to comparing the performance of musicians and non-musicians in basic auditory perception tasks. Pitch is a fundamental dimension of auditory perception, which plays an essential role in most forms of music; differences in pitch define musical intervals, and pitch variations over time are used to convey melodies. Accordingly, one might expect musicians to show substantially enhanced performance, compared to non-musicians, in tasks that involve discriminating sounds along the pitch dimension. This expectation is partly confirmed by the results of two earlier studies (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001). These results demonstrate significantly smaller frequency discrimination thresholds, i.e., higher performance, in musicians than in non-musicians. However, the difference between the two groups is relatively modest (a factor of about two, on average), and the results of Kishon-Rabin et al. (2001) suggest that, within an hour of practice of the frequency discrimination task, non-musicians with no prior psychoacoustical training can obtain thresholds that are, on average, as small as those achieved at first by experienced musicians.

* Corresponding author. Tel.: +1 617 324 0476; fax: +1 617 258 7003.
E-mail address: cmicheyl@mit.edu (C. Micheyl).

0378-5955/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2006.05.004

Given the extensive and rigorous training undertaken by many musicians, it may seem surprising that the initial difference in frequency discrimination thresholds between musicians and non-musicians is not larger than just a factor of two, and that an hour or less of practice in the frequency discrimination task is sufficient for non-musicians to obtain thresholds as small as the musicians. In fact, several observations suggest that the difference between the musicians' and the non-musicians' average thresholds observed in the two studies of Spiegel and Watson (1984) and Kishon-Rabin et al. (2001) may not represent the full benefit of musical training on pitch discrimination performance. Firstly, some of the "non-musicians" in Spiegel and Watson (1984) study had a "high degree of musical or psychoacoustical experience" (Spiegel and Watson, 1984, p. 1692), which may have contributed to blurring the difference between the two groups. Secondly, more than half of the musicians in Kishon-Rabin et al. (2001) study had a contemporary (i.e., modern or jazz) background, and these musicians exhibited significantly higher frequency discrimination thresholds than the seven other musicians, who had a classical music background; retrospectively, this suggests that the average threshold of the musicians in that study does not represent the optimal performance possible through musical training. Thirdly, in both Spiegel and Watson's and Kishon-Rabin et al.'s studies, the subjects were given relatively limited opportunity to practice the frequency discrimination task and, even at the end of the test session, their average thresholds were elevated compared to those reported in the psychoacoustical literature using highly trained listeners (e.g., Moore, 1973; Wier et al., 1977). While these elevated thresholds may reflect a genuine limitation on the sensory discrimination ability of untrained listeners, they could also reflect the listeners' incomplete familiarization with the specifics of the psychoacoustical test and/or with the unnatural sounds used in it; the latter factor may have prevented the musicians from performing as well as they could have, had they been given more time to acquaint themselves with the procedure and stimuli.[1]

In the present study, we attempted to avoid the potential limitations of earlier studies, and to maximize group separation. This was done in two ways. Firstly, we included in the group of musicians only full-time classical musicians who had 10 years or more of experience playing an instrument, and we included only individuals who had no musical training in the group of non-musicians. We also excluded subjects with any prior psychoacoustical experience from either group, so that all listeners were on equal footing in this respect. Secondly, we let the musicians practice the psychoacoustical task long enough to ensure that their thresholds had reached an asymptotic value, thus allowing comparisons of performance at a point where incomplete familiarization with the procedure or the stimuli was no longer a potential confound. In a further effort to optimize the expected benefit of musical training, we used both pure and harmonic complex tones, and selected the reference frequency of these tones to correspond to a note on the musical scale.

One last feature of the present study, which also distinguishes it from earlier studies, relates to the fact that we systematically tested both the right and the left ears in each listener using monaural, rather than diotic test tones, and also used contralateral noise masking, in order to test for ear differences in pitch discrimination performance within or across the two study groups. The results of several psychoacoustical, neuropsychological, and brain-imaging studies suggest the existence of hemispheric and ear asymmetries in various tasks involving pitch perception (Bever and Chiarello, 1974; Divenyi et al., 1977; Johnson, 1977; Johnsrude et al., 2000; Messerli et al., 1995; Peretz and Morais, 1983; Peretz and Babaï, 1992; Sidtis, 1980, 1981; Zatorre et al., 1992). In most cases, the asymmetry was found to be in favor of the right hemisphere. However, some investigators have reported opposite patterns of ear dominance in musicians and non-musicians, perhaps reflecting differences in listening strategies (Bever and Chiarello, 1974; Johnson, 1977; Messerli et al., 1995).

Using the above inclusion criteria and test conditions, two experiments were performed. The first experiment involved 30 musicians and 30 non-musicians, and enough trials to ensure that the musicians' thresholds reached an asymptotic value before the end of the test session. However, because the number of trials in experiment 1 was not sufficient for the non-musicians to achieve an optimal level of performance, a second experiment was performed, which involved protracted training in eight additional non-musician listeners, using one of the stimulus conditions from experiment 1. Although several earlier studies have documented long-term training effects in frequency discrimination in non-musicians (Amitay et al., 2005; Ari-Even Roth et al., 2003, 2004; Campbell and Small, 1963; Delhommeau et al., 2002, 2005; Demany, 1985; Demany and Semal, 2002; Grimault et al., 2002, 2003; Irvine et al., 2000; Wright and Fitzgerald, 2005), we reasoned that documenting learning effects in non-musicians using the same stimuli and test procedure as in one of the conditions

[1] The notion that the listeners' initial lack of familiarity with the procedure and/or stimuli could lead to an under-estimation of the actual difference in sensory discrimination abilities between musicians and non-musicians can be understood in terms of a signal-detection-theoretic model (Green and Swets, 1966) wherein frequency discrimination performance is limited by two types of additive sources of internal noise: "sensory" noise, which imposes an absolute upper limit on frequency discrimination abilities and is smaller in musicians than in non-musicians, and "cognitive" noise, which reflects the listeners' lack of familiarity with the specifics of the procedure and stimuli, and is the same for musicians and non-musicians. Under this model, the mean frequency discrimination threshold of the musicians can be expressed as $\theta_m \propto \sqrt{s_m^2 + c^2}$, and that of the non-musicians as $\theta_n \propto \sqrt{s_n^2 + c^2}$, where s^2 and c^2 denote the variance of the sensory and cognitive noises, respectively, the subscripts m and n refer to musicians and non-musicians, respectively, and $\propto$ is used to indicate a proportionality relationship. Based on these equations, the musicians-to-non-musicians threshold ratio, θ_m/θ_n, decreases as the variance of the "cognitive" noise, c^2, increases.

of experiment 1 would permit direct comparisons between the two experiments, and the two groups of listeners.

2. Methods

2.1. Listeners

This study involved a total of 68 subjects who had normal hearing (as defined by pure-tone absolute thresholds of 20 dB HL or less at octave frequencies between 500 and 8000 Hz), no history of hearing disorders, and no prior psychoacoustical experience. Sixty subjects took part in the first experiment. Half of them were either music students from the National Superior Music Conservatory of Paris, or professional musicians from the same Music Conservatory's symphony orchestra (Orchestre des Lauréats du Conservatoire). Recruiting the musicians from a National Music Conservatory ensured that music was a full-time activity for these individuals. All had been playing their instrument for over 10 years, and they were still practicing music on a daily basis (for several hours a day) at the time of the study. Specifically, 18 of the musicians had been playing an instrument for 12–16 years, 6 for over 16 years, and the remaining 6 for less than 12 years; the mean number of years of instrumental practice across the whole group was 14 years. One third of the musicians were introduced to music before the age of 6, 12 between the ages of 6 and 8, and the remaining 8, after 8. One third of the musicians played a keyboard instrument (the piano). Eleven other musicians played a wind instrument (the flute, the clarinet, or the trumpet). The remaining 9 played a string instrument (mainly, the violin or the viola). Seventeen of the musicians reported having absolute pitch, although this was not formally verified in this study.

The other 30 subjects in experiment 1 had never learned to play any musical instrument. The musicians and non-musicians were similar in age (mean = 22 years, SD = 1.9 for the musicians; mean = 20 years, SD = 1.29 for the non-musicians), gender (15 female in each group), handedness (only right-handed subjects were included; mean score at the Edinburgh Handedness Inventory, Oldfield, 1971 = 70%, SD = 27 for the musicians and 66%, SD = 18 for the non-musicians) and education (all subjects had a college-level education).

Eight additional non-musician listeners took part in the second experiment. They were aged between 18 and 21 years (mean = 20 years, SD = 1.76). None of these listeners had taken part in the first experiment.

2.2. Stimuli

Two types of test tones were used in this study: pure tones and harmonic complex tones. The latter were obtained by summing four sinusoids with frequencies corresponding to harmonics 2 to 5 of a fundamental frequency (F0) of 330 or $330 + \Delta f$ Hz. The four sinusoids all started in sine (0°) phase and had equal amplitudes. The pure tones consisted of a single sinusoid with a frequency of 330 or $330 + \Delta f$ Hz. The 330 Hz frequency was chosen because it corresponds to the E4 note on the equal-tempered Western musical scale with which the musicians were highly familiar.

All tones were 200 ms each in duration, including 20-ms raised-cosine on and off ramps, and were presented monaurally under headphones at a level of 65 dB SPL. Depending on the condition being tested, the contralateral ear was either not stimulated (monaural condition) or stimulated with noise bursts (contralateral-noise condition), which were synchronous with the tones. The synchronous contralateral noise masker was introduced in an attempt to promote inter-hemispheric competition (Kimura, 1964) and prevent binaural neurons activated primarily by the contralateral ear from contributing substantially to performance. Two contralateral-noise conditions were tested: in the contra/on-frequency condition, the noise was lowpass filtered at 4 kHz, so that its frequency range encompassed that of the test tones (i.e., from 330 Hz up to about 2 kHz); in the contra/off-frequency condition, the noise was bandpass filtered between 4 and 8 kHz, so that its spectrum did not overlap with that of the test tones. The contra/off-frequency condition was used to check whether the influence of the contralateral noise was merely due to the presence of noise in the opposite ear (as would be the case if the contralateral noise had merely a distracting effect), or if the noise had to occupy the same frequency region as the tones in order to significantly influence performance (which would be more consistent with the hypothesis that the noise influenced frequency-specific neural populations). The overall level of the noise was approximately 85 dB SPL. It can be estimated that the RMS level of the noise passing inside auditory filters with center frequencies corresponding to the tones presented in the other ear was between about 4.8 and 10.8 dB higher than the RMS of the tones. Thus, had the noise been presented in the same ear as the tones, it would have masked them. However, because the noise was presented in the contralateral ear, the tones were always audible.

2.3. Procedure

A two-interval, two-alternative forced-choice procedure was used to measure frequency and F0 discrimination thresholds. On each trial, two tones were presented successively. One of the tones (the standard) had a frequency (or F0) of 330 Hz. The other tone (the target) had a frequency or F0 of $330 + \Delta f$ Hz. The order of presentation of the two tones was random, with the higher frequency (or F0) tone being equally likely *a priori* to fall in the first or the second interval. The listener's task was to indicate in which interval the higher frequency tone was presented. Visual feedback was provided following each response. The frequency difference between the two tones, Δf, was varied adaptively using a two-down/one-up rule, which tracked the 70.7%-correct point on the psychometric function. At the beginning of a "run" of the adaptive procedure, Δf was set to 20% of 330 Hz (i.e., 66 Hz), a value large enough to ensure that the listeners would have no difficulty perceiving the difference between the two tones. Following two consecutive correct responses, Δf was decreased by a certain factor; it was increased by the same factor after any incorrect response. Up to the fourth reversal in the direction of the change in the value of Δf, the factor by which Δf was decreased or increased was equal to 2; following the fourth reversal, it was set to $\sqrt{2}$. On the twelfth reversal, the tracking procedure stopped and the discrimination threshold was computed as the geometric mean of the Δf values corresponding to the last 8 reversals. The resulting values were expressed as percentages of the standard frequency or F0.

2.4. Experiment 1 design

Experiment 1 involved a single test session, which lasted between two and a half and 3 h per subject, including breaks. During the session, the listener's frequency discrimination thresholds were measured five times in each of twelve different stimulation conditions, which resulted from the combination of tone type (pure or complex), test ear (right or left), and presentation mode (monaural, contra/on-frequency noise, and contra/off-frequency noise). The conditions were tested in pseudo-random order: all 12 conditions were tested in one block, then another block of the same 12 conditions was tested, and so forth, until five blocks were completed. The order of testing of the conditions within each block was completely randomized, and could differ across blocks as well as across listeners. As a result, each listener in experiment 1 performed a total of 60 runs of the adaptive threshold-tracking procedure. Typically, a run involved around 40 trials (the actual number of trials could vary across runs because the turnpoints in the adaptive staircase did not always occur at the same point). Accordingly, each listener from experiment 1 performed around 2400 trials in total.

2.5. Experiment 2 design

Experiment 2 involved seven test sessions, which took place on different days and lasted between 1.5 and 2.5 h each. The first and the last sessions are referred to as the pre- and post-training sessions. During these two sessions, the listener's pure-tone frequency thresholds were measured in four different test conditions, which resulted from the combination of test ear (right or left) and presence or absence of contralateral noise – similar to the on-frequency noise used in experiment 1. On each of the two sessions, the four conditions were tested in a pseudo-randomized blocked fashion (similar to experiment 1). Seven blocks were consecutively performed, yielding a total of 28 runs (an estimated 1120 trials) per session. On each of the five training sessions that followed the multi-condition pre-training session, the listeners performed 30 runs in a single test condition, with the pure tones presented monaurally to the right ear and no contralateral noise. These five mono-condition training sessions are hereafter referred to as the training sessions proper and numbered from 1 to 5. Thus, over the course of the mono-condition training period, each listener performed a total of 150 runs, representing an estimated 6000 trials.

2.6. Apparatus

Stimuli were generated in the time domain on a Pentium computer, played out at a sampling rate of 44.1 kHz via a 16-bit Roland UA30 sound card, and delivered through Sennheiser HD465 headphones. Listeners were tested in a sound-insulated room.

3. Results

3.1. Experiment 1: differences between musicians and non-musicians

The mean frequency and F0 discrimination thresholds (FDTs and F0DTs, respectively) measured in the 30 musicians and the 30 non-musicians of experiment 1 are shown in Fig. 1. On average across all test conditions and runs, the discrimination thresholds of the musicians (0.13%) were more than six times smaller than those of the non-musicians (0.86%). The difference proved highly significant when tested using an ANOVA on the log-transformed thresholds[2] [$F(1, 58) = 77.11$, $p < 0.001$]. Further inspection of Fig. 1 reveals the following. First, the musicians' advantage was more pronounced with complex tones than with pure tones. This was confirmed by a significant group × tone type interaction [$F(1, 58) = 11.97$, $p = 0.001$]. Second, while contralateral noise was found to cause a significant increase in thresholds [$F(2, 116) = 19.57$, $p < 0.0001$], this effect was mainly due to the contra/on-frequency noise [$p < 0.0005$]; the effect of the contra/off-frequency noise failed to reach statistical significance [$p = 0.058$]. Furthermore, the influence of the contralateral noise did not differ significantly between the musicians and the non-musicians [$F(2, 116) = 2.60$, $p = 0.078$]. Finally, there were no consistent differences in thresholds between the right and left ears [$F(1, 58) = 2.3$, $p = 0.13$]. Marginally, a significant left-ear

[2] The application of a logarithmic transformation on the thresholds prior to their submission to parametric statistical analyses was motivated by theoretical and practical considerations. From a theoretical standpoint, it is consistent with the notion that the underlying perceptual scale in frequency discrimination tasks is a ratio scale. From a practical standpoint, the logarithmic transformation is consistent with the use of multiplicative (rather than additive) steps in the adaptive threshold-tracking procedure. Furthermore, the logarithmic transformation helps to correct for the fact that the variability of frequency discrimination thresholds increases with their magnitude, which might otherwise result in a violation of the homoscedasticity assumption of the parametric statistical tests. Note that the use of a logarithmic or square-root transformation of frequency discrimination thresholds is common practice in the psychoacoustic literature (e.g., Irvine et al., 2000; Delhommeau et al., 2002, 2005; Demany and Semal, 2002).

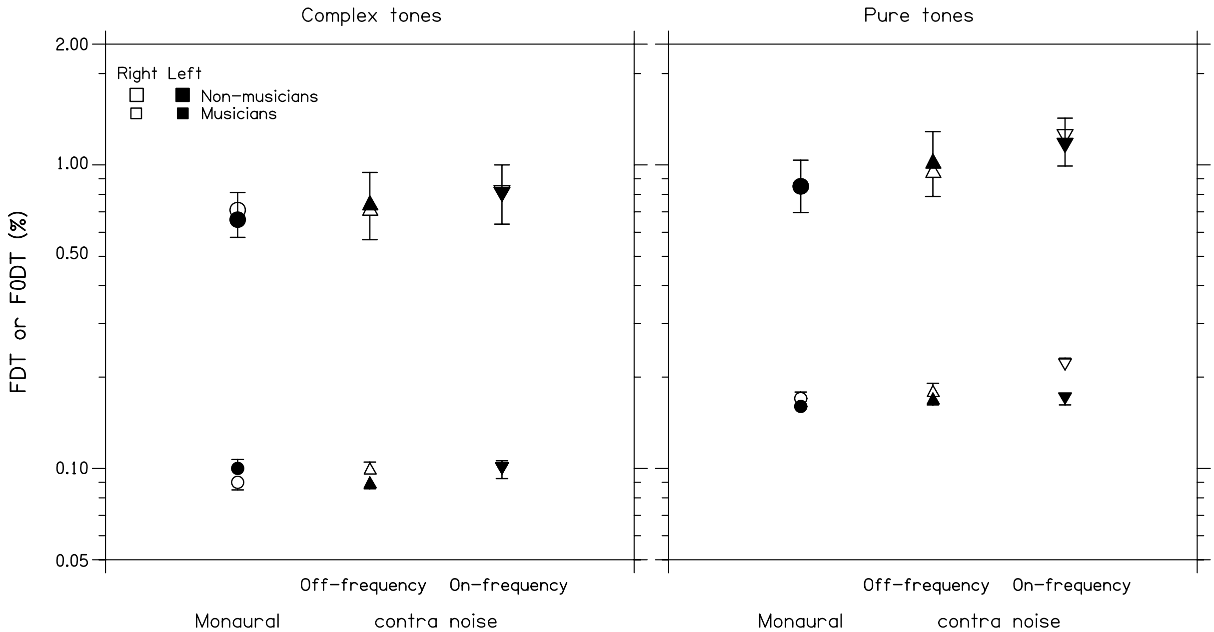

Fig. 1. Mean frequency- and F0-discrimination thresholds (FDTs and F0DTs, respectively) in the musicians and the non-musicians. Each data point in this figure corresponds to the geometric mean, across all 30 listeners from a given group, of the five successive threshold measurements in a given test condition. The left-hand panel shows thresholds measured using complex tones (F0DTs); the right-hand panel, thresholds measured using pure tones (FDTs). Different symbols are used to denote thresholds measured using different modes of stimulus presentation, indicated underneath the x-axis: monaural (circles), with off-frequency contralateral noise (upward-pointing triangles), with on-frequency contralateral noise (downward-pointing triangles). As indicated by the inset legend, empty symbols denote thresholds measured with the test tones in the right ear while solid symbols denote thresholds measured in the left ear, and larger symbols are used to indicate data from the non-musicians. All thresholds are expressed as percentages of the standard frequency or F0. The error bars show plus or minus one standard error around the corresponding geometric mean. Overlapping error bars are not displayed. Note that, for the musicians, the standard errors of the means were sometimes too small for the corresponding error bars to be visible at this scale.

advantage was found in the condition involving pure tones and contralateral on-frequency noise, for the musicians only [$F(1,29) = 22.16$, $p < 0.001$].

Fig. 2 shows individual data for each group. Each data point in this figure corresponds to the mean threshold of one listener in the monaural stimulation condition, after

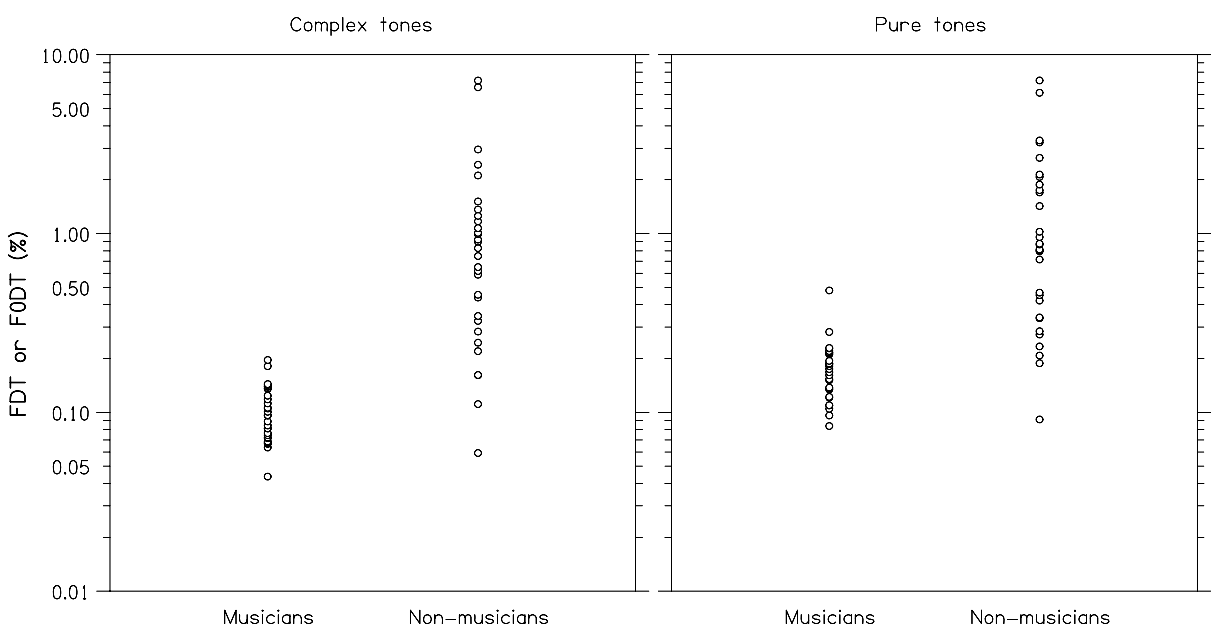

Fig. 2. Individual FDTs and F0DTs in musicians and non-musicians. Each data point in this figure corresponds to an individual listener, and was computed as the geometric mean of the 10 threshold measurements obtained in the monaural stimulation condition, after pooling data from the right and left ears. As in the previous figure, the left-hand panel shows F0DTs, measured using complex tones, while the right-hand panel shows FDTs, measured using pure tones. Within each panel, the data from the musicians are shown on the left; those of the non-musicians, on the right.

the data were averaged across the two ears and the five consecutive measures. As in the previous figure, F0DTs (measured with complex tones) are shown in the left-hand panel and FDTs (measured with pure tones) are shown in the right-hand panel. This plot makes apparent the large variability of results in the non-musicians' group, with some of the non-musicians having smaller thresholds than some of the musicians, while others had thresholds more than an order of magnitude larger than the musicians. For pure tones, nearly half the non-musicians had thresholds in the same range as the musicians; for complex tones, the proportion of non-musicians whose thresholds were in same range as the musicians was less than a third.

Fig. 3 shows how discrimination thresholds improved across trial runs in the two subject groups. Although a significant overall improvement in thresholds was observed in both groups [$F(4,232) = 50.35$, $p < 0.001$], the improvement was markedly larger in non-musicians than in the musicians [block × group interaction: $F(4,232) = 15.15$, $p < 0.001$]. The non-musicians improved by a factor of approximately 3 on average, between the first and the third blocks; the musicians only improved by a factor of about 1.33. Furthermore, in the musicians, post-hoc comparisons (Tukey's HSD) demonstrated a significant improvement ($p < 0.001$) between the first and second blocks only, and not thereafter ($p > 0.5$ for all further pairs). For the non-musicians, significant improvements were observed until the third block ($p < 0.001$ for block 1 vs. block 2, $p < 0.05$ for block 2 vs. block 3, and $p > 0.05$ thereafter), and visual inspection of the data indicates a trend for thresholds to continue improving throughout the five blocks in this group, suggesting that the absence of significant difference beyond the third block might be due to insufficient statistical power. On the first block, the thresholds of the non-musicians were an order of magnitude larger than those of the musicians. When considering only the last two blocks, which the listeners performed after they had been practicing the task for more than an hour (including rest times) and completed 36 runs (an estimated 1140 trials), thresholds were still approximately five times smaller in the musicians than in the non-musicians, and the difference was still highly significant [$F(1,58) = 58.19$, $p < 0.001$].

Fig. 4 shows the mean frequency discrimination thresholds measured in three sub-groups of musicians, based on the family of musical instrument played. On average, musicians who played a keyboard instrument (mainly, the piano) had larger thresholds than those who played other families of instruments (strings and winds) [contrast analysis: $F(1,27) = 5.98$, $p < 0.05$]. This effect was independent of the type of stimulus used; it was observed with the pure tones [$F(1,27) = 5.29$, $p < 0.05$] as well as with the complex tones [$F(1,27) = 4.35$, $p < 0.05$]. Despite their worse performance compared to other musicians, the musicians in the keyboard group still had significantly lower thresholds on average than the non-musicians [$F(1,38) = 21.94$, $p < 0.001$].

No significant correlation was found between the thresholds measured in the musicians and either the number of years that they had practiced music ($r = 0.114$, $p = 0.55$) or the age at which they started practicing ($r = 0.005$, $p = 0.98$). Similarly, no significant difference in thresholds was found between musicians who claimed to possess absolute pitch and those who did not [$F(1,28) = 0.18$, $p = 0.67$].

3.2. Experiment 2: influence of psychoacoustical training in non-musicians

The results of experiment 2 are illustrated in Fig. 5. This figure shows how the thresholds of the eight non-musicians improved across the seven sessions of that experiment. For the pre- and post-training sessions, the different symbols represent the mean thresholds measured in different testing conditions (right or left ear, monaural or dichotic), as indicated in the legend of Fig. 5. For these two sessions, each symbol represents the (geometric) mean of 56 threshold measurements, which resulted from pooling data across listeners and blocks (8 listeners × 7 blocks on each session). The data of the five specific training sessions (numbered 1–5) are indicated by empty circles, the symbol that was also used in Figs. 1 and 3 to represent data obtained in the monaural right-ear testing condition. This was the only stimulus condition used during the five specific training sessions of the current experiment. For these sessions, each symbol represents the geometric mean of 240 threshold measurements, which resulted from pooling data across listeners and runs (8 listeners × 30 runs per session). An ANOVA on the log-transformed thresholds of the mono-condition training sessions (i.e., sessions 1–5) revealed a significant improvement between the first two such sessions [$F(1,7) = 7.69$, $p = 0.028$]. Despite an apparent trend for thresholds to improve between sessions 2 and 3, the improvement failed to reach statistical significance [$F(1,7) = 2.05$, $p = 0.195$]. Detailed inspection of the individual data revealed that this was because the thresholds of one subject increasing markedly between the second and third sessions. The reasons for this increase remain unclear; it could be due to the subject being unusually tired or unmotivated at the time of the third session. However, evidence that learning continued beyond the second training session is provided by the finding of a significant difference in average thresholds between the second and fourth sessions [$F(1,7) = 6.437$, $p = 0.040$].

By the end of the training period, the mean pure-tone frequency threshold of the non-musicians in this experiment was as small as, and not statistically different from, that measured in the musicians of the previous experiment – approximately, 0.16%, as illustrated by the horizontal dashed line (Student's *t*-test on independent samples: $p = 0.18$). This lack of significant difference cannot be ascribed simply to insufficient statistical power due to a smaller sample size here than in experiment 1, because there was a highly significant difference ($p < 0.001$) between the thresholds measured on the first training session of the

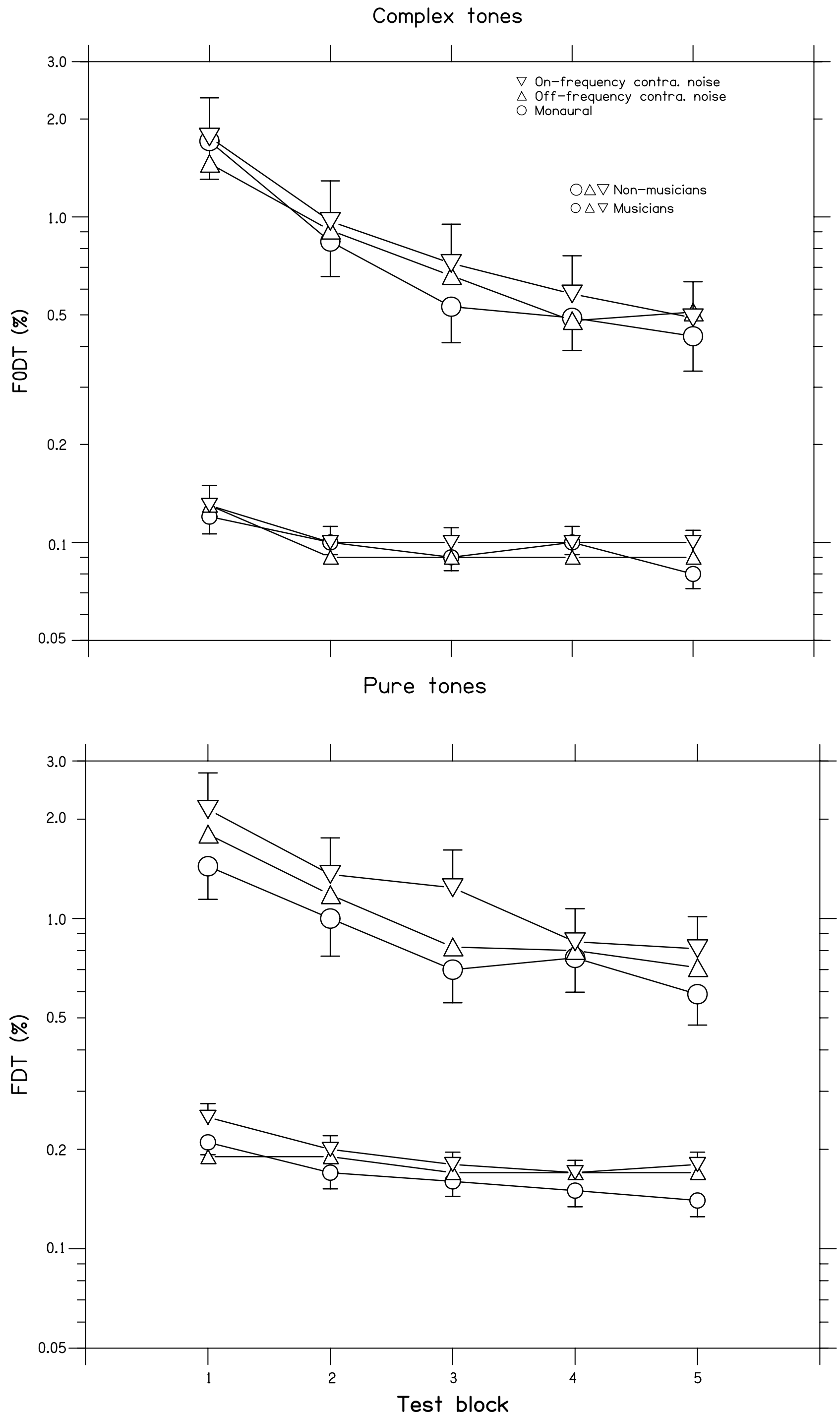

Fig. 3. Influence of short-term practice on FDTs and F0DTs in the musicians and the non-musicians. Each data point in this figure corresponds to the geometric mean, across all 30 listeners from a given group, of the thresholds measured in the left and right ears using a given mode of stimulus presentation, on a given block of measurements. The five consecutive test blocks are indicated by the numbers underneath the *x*-axis of the bottom plot. The upper panel shows F0DTs (complexes tones); the lower panel shows FDTs (pure tones). Thresholds measured using different modes of stimulus presentation are denoted by different symbols, as indicated by the legend; these symbols are the same as in Fig. 1. Also similar to Fig. 1, the data of the non-musicians are indicated by larger symbols than those of the musicians. The error bars show plus or minus one standard error around the corresponding geometric mean. Overlapping error bars are not displayed.

current experiment and the pure-tone thresholds of the musicians from experiment 1.

In order to determine at what point of the training phase the thresholds of the non-musicians in the current experiment approached those of the musicians, the mean threshold measured on each of the five specific training sessions in the current experiment was compared with the (log-transformed, geometric) mean pure-tone monaural threshold measured in the musicians on the last three blocks of runs of experiment 1. The results revealed that on the first

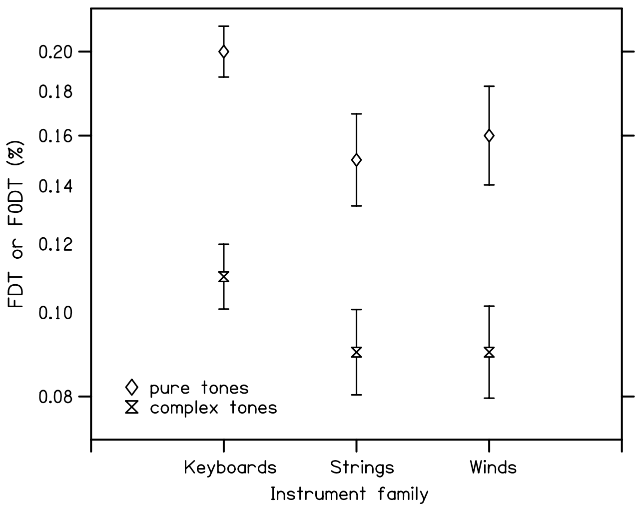

Fig. 4. Mean FDTs and F0DTs in three sub-group of musicians, sorted based on the family of their primary instrument. The names of the three instrument families are indicated underneath the x-axis. Each data point in this figure corresponds to the geometric mean, across all musicians who played an instrument from the considered family, of the FDTs or F0DTs measured in the left and right ears, under monaural testing conditions. FDTs are indicated by diamonds; F0DTs, by hourglass-like symbols. The error bars show plus or minus one standard error around the corresponding geometric mean. Overlapping error bars are not displayed.

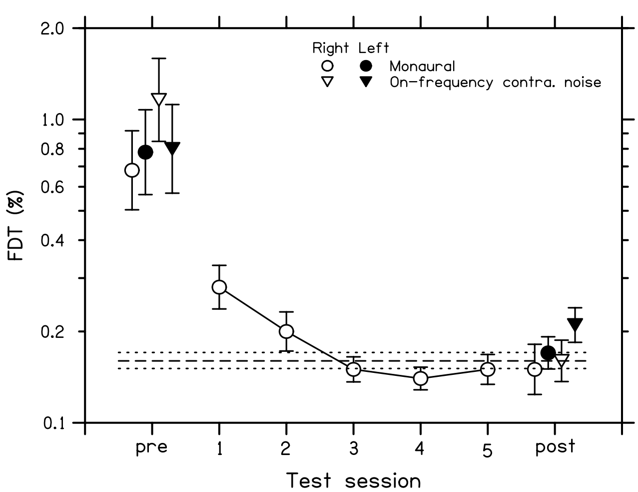

Fig. 5. Influence of protracted psychoacoustical training on FDTs in non-musicians. Each data point corresponds to the geometric mean, across the eight non-musician listeners who took part in experiment 2, of the FDTs measured in a given test condition, on a given test session. For the pre- and post-training sessions, different symbols are used to denote thresholds measured in the left and right ears, in the absence or in the presence of contralateral on-frequency noise, as indicated by the legend. The data points were slightly shifted away from each other horizontally in order to avoid clutter. The mean FDTs from the five specific training sessions (monaural presentation mode, right ear only), numbered from 1 to 5 on x-axis, are indicated by empty circles. The dashed line shows the geometric mean FDT measured in the same (i.e., monaural, right ear) condition in the musicians from experiment 1. The surrounding dotted lines show plus and minus one standard error around this mean.

specific training session, the non-musicians' thresholds were still significantly higher than the musicians' ($t = 3.29$, $p < 0.005$). On the second training session, the difference was already no longer significant, but a trend was present ($p < 0.085$). On subsequent sessions, the difference was clearly non-significant ($p \gg 0.05$). Based on these results, it can be concluded that between 4 and 8 h of psychoacoustical training were required for the non-musicians to obtain pure-tone frequency discrimination thresholds as small as the musicians, on average – note that this estimate includes the two-hour, multi-condition pre-training session. Interestingly, although the non-musicians were allowed further training, their thresholds did not become significantly smaller than those of the musicians.

4. Discussion

4.1. How large is the musicians' advantage in pitch discrimination?

The results of experiment 1 provide further evidence for enhanced pitch discrimination performance in musicians, compared to non-musicians. This is in line with earlier results from Spiegel and Watson (1984) and Kishon-Rabin et al. (2001). However, whereas the musicians' advantage in these earlier studies amounted to a factor of about two on average, here, the difference in frequency discrimination thresholds between the musicians and the non-musicians was initially found to correspond to a factor of about six. Even after the listeners had practiced the pitch discrimination task for approximately 2 h and performed an estimated 1440 trials, the non-musicians' thresholds were still around four times larger, on average, than those of the musicians.

The difference in effect size between the present and earlier studies is unlikely to be due merely to the use of different testing conditions, because the stimuli and procedure used here were not dramatically different from those used in earlier studies. Of course, the nominal stimulus frequency used here (330 Hz) was not exactly the same as in previous studies; however, Spiegel and Watson (1984) and Kishon-Rabin et al. (2001) used frequencies that were not too far off on either side (e.g., 250 and 430 Hz), which makes it highly unlikely that the difference in effect size between both of these studies and the current one is due to some frequency-specific factor. Furthermore, Spiegel and Watson (1984) tested both musical and non-musical frequencies, and they found the difference in thresholds between musicians and non-musicians to be approximately the same in the two cases; this makes it unlikely that the difference in effect size between the current and previous studies is due primarily to our choice of testing at a frequency that corresponds exactly to a musical note on the diatonic scale.

The most likely explanation for the finding of a larger difference in pitch discrimination thresholds between musicians and non-musicians in the present study than in earlier studies relates to our use of more stringent selection criteria. Specifically, our decision to include in the musician group only individuals who had a classical music background and had played an instrument for over 10 years, probably contributed to the musicians of the present study having smaller frequency discrimination thresholds, on average, than that of Kishon-Rabin et al. (2001) – more than half of which was comprised of jazz or modern musicians. Consistent with this, Kishon-Rabin et al. (2001) found that the classical musicians in their study sample had lower frequency discrimination thresholds than the modern-style musicians. A possible explanation for the difference in pitch discrimination performance between these two sub-groups of musicians is that classical music places more emphasis on correct tuning than other musical styles, and that this emphasis on correct tuning promotes the development of more accurate pitch discrimination abilities. Along the same lines, our decision to include in the non-musician group only individuals who had no musical training and no prior psychoacoustical experience may also have contributed to the separation between musicians and non-musicians being larger here than in the Spiegel and Watson's (1984) study. Indeed, in that study, the control group included subjects with a "high degree of musical or psychoacoustical experience", and the authors pointed out that these subjects were more likely to have thresholds in the same range as the musicians.

Thus, the most likely reason for the identification of large differences in pitch discrimination performance between musicians and non-musicians in the present study was the use of stringent selection criteria for the two groups. One implication of the present finding of larger differences in pitch discrimination thresholds between musicians and non-musicians is that classical musical training can have a more profound influence on initial pitch discrimination performance than suggested by the results of earlier studies.

4.2. Can musical training alone lead to optimal pitch discrimination performance?

In this study, as well as those of Spiegel and Watson (1984) and Kishon-Rabin et al. (2001), both non-musicians and musicians showed some improvement in thresholds with practice. In all three studies, the improvement was found to be much less marked for the musicians than for the non-musicians. A question that remained unanswered by previous studies was how long musicians have to train in order to obtain optimal thresholds. At the end of Kishon-Rabin et al.'s (2001) brief experiment (1 h per subject), the thresholds of the musician's group were still markedly larger (by a factor of 4 or more) than those reported in the psychoacoustic literature for highly trained listeners tested under comparable conditions (e.g., Moore, 1973). Thus, whether sensory or procedural, the learning was obviously not complete in that group. Spiegel and Watson's (1984) data are more difficult to compare to other data in the literature, due to the use of unusual testing conditions (i.e., tape-recorded stimuli presented via loudspeakers) and of an unusual threshold-estimation procedure (i.e., visual fitting of a line through the listener's psychometric function). Although these authors pointed out that their musicians' thresholds were similar to those reported in studies using highly trained listeners tested under headphones, in fact, careful inspection reveals that the musicians' thresholds in that study were larger than those usually observed in highly trained listeners; it remains unclear whether this difference is due solely to the use of loudspeakers rather than headphones, or to incomplete learning in the musicians.

In contrast, in the present study, the asymptotic average value of the musicians' pure-tone frequency discrimination thresholds (around 0.15%) was similar to those reported in studies involving highly trained listeners: between 0.26% at 250 Hz and 0.15% at 500 Hz for the "best" listener in Moore (1973), between 0.5% at 200 Hz and 0.25% at 400 Hz for the mean across listeners in Wier et al. (1977) study. Thus, the present results reveal that classical musicians can achieve optimal frequency discrimination performance within a relatively short time, and that they need little psychoacoustical training. In fact, on the first block (which was completed within the first hour of testing) the musicians from this study already had average thresholds lower than 0.2%, and no further improvement was observed following the second block.

Considering that the rapid initial improvement in the musicians' thresholds observed here and in previous studies may reflect these listeners' necessary adaptation to the attentional demands of the test or other forms of procedural learning, rather than a genuine improvement in sensory discrimination abilities, the data do not rule out the possibility that musical training does in fact lead to the development of optimal pitch discrimination performance. Unfortunately, neither the present results nor those of earlier studies permit a clear conclusion regarding the nature (procedural or sensory) of the perceptual learning effects observed in musicians; further experimentation is required to clarify this point.

4.3. How long must non-musicians train?

Kishon-Rabin et al. (2001) found that non-musicians had thresholds as small as those obtained initially (i.e., on the first block of runs) by musicians after performing only four runs of an adaptive threshold-tracking procedure at each of three test frequencies (i.e., an hour or less of practice). The results of experiment 2 indicate that between 4 and 8 h of practice in the frequency discrimination task were needed, on average, for the non-musicians to obtain thresholds as small as those of the musicians. One way to reconcile the present results with this earlier finding relates to the possibility, which we mentioned in the Introduction, that thresholds were initially elevated because of the listeners' lack of familiarity with the psychoacoustical test procedure and stimuli. From that point of view, the benefit of musical training may not fully reveal itself on initial runs because, at this stage, thresholds may be determined in large part by factors that are unrelated to the subject's true sensory discrimination abilities.

Another possible explanation for the different outcome between Kishon-Rabin et al. (2001) study and the present one relates to differences in inclusion criteria, and their influence on group separation. Our decision to include only classical musicians may explain the finding of a substantially larger initial difference in thresholds between the two groups here than in Kishon-Rabin et al. (2001) study. In turn, the larger initial separation between the two groups can explain the longer time required for the thresholds of the non-musicians to become statistically indistinguishable from those of the musicians.

In comparing the non-musicians' learning data from experiment 2 and the musicians' data from experiment 1, we have so far ignored the fact that the former were collected in the context of mono-condition test sessions while the latter were obtained in a multi-condition context. Thus, it could be argued that the comparison is biased, because the testing of multiple conditions in random order may have limited listeners' ability to focus on each condition, leading to worse performance. While this remains a possibility, the observation that the musicians' thresholds, measured during multi-condition testing, were as small as those measured in highly trained listeners in other studies (e.g., Moore, 1973; Wier et al., 1977) suggests that multi-condition testing had no major detrimental impact on performance.

4.4. Which type of musical instrument is practiced matters

Another finding of the present study relates to the poorer pitch discrimination performance of pianists, compared to the other classical musicians who played wind or string instruments. A possible explanation for this observation is that, whereas musicians who play string or wind instruments usually tune their instrument at the beginning of each practice or performance, pianists usually do not tune their instrument themselves – tuning a piano is regarded by most musicians as a difficult exercise, which requires the intervention of a specially trained professional. It is conceivable that self-tuning of one's musical instrument promotes the development of finer pitch discrimination abilities. A similar explanation was offered by Spiegel and Watson (1984) to explain their observation of smaller frequency discrimination thresholds in musicians whose primary instrument was string, woodwind, or brass, compared to musicians who played other types of instruments and did not tune their instrument themselves or used electronic tuners. However, because of the small size of the latter sub-group (three subjects), Spiegel and Watson's conclusion on this point remained tentative. In contrast, one third of the musicians from the present study (i.e., 10 subjects) had the piano as their primary instrument, allowing a firmer conclusion to be drawn.

Kishon-Rabin et al. (2001) reported not finding evidence in their data to support the hypothesis that musicians who tune their instruments are better able to discriminate small pitch changes. They suggested that frequency discrimination performance was more dependent upon musical genre (classical vs. contemporary) than on the type of instrument played *per se*. However, careful inspection of Table 1 in their article indicates that while all the classical musicians in that study played self-tunable instruments (the violin, the viola, the bassoon, or the French horn), slightly more than half of the contemporary musicians played either percussion (the tuning of which is, arguably, more rudimentary than that of other instruments, because of the impulsive and broadband nature of the percussion sounds) or keyboards (which are usually not self-tuned). From that point of view, Kishon-Rabin et al.'s data are not inconsistent with the hypothesis that the tuning of one's musical instrument promotes pitch discrimination abilities.

4.5. A larger advantage of musicians with complex tones

Another interesting result from the present study corresponds to the finding of a larger pitch discrimination advantage of musicians over non-musicians for complex tones than for pure tones. A possible explanation of this finding relates to the observation that, since most natural musical sounds are harmonic complexes rather than pure tones, musicians are mainly (if not exclusively) exposed to harmonic complex tones. It is conceivable that this repeated exposure to complex tones promotes the development of enhanced pitch discrimination for such tones. A similar argument has been invoked by other authors to explain the observation that absolute pitch possessors display higher pitch identification performance when they are tested with complex tones than when pure tones are used (Lockhead and Byrd, 1981; Miyazaki, 1989).

On the other hand, the present results clearly demonstrate that the superiority of musicians over non-musicians for pitch discrimination extends to pure tones. This finding is consistent with the results of Spiegel and Watson (1984), who also found smaller thresholds in musicians than in non-musicians for both pure and complex tones – although it is worth noting that the complex tones in that study only contained odd harmonics. The finding that the musicians' advantage extends to pure tones, despite musical sounds being essentially harmonic complexes, may be explained in the light of data in the literature, which indicate that the improvements in pitch discrimination performance that are induced by practice with exclusively complex tones generalize partly to pure-tone pitch discrimination (Demany and Semal, 2002). The finding that this generalization of learning is incomplete (Demany and Semal, 2002) is consistent with our finding of a larger musicians' advantage with harmonic complex tones than with pure tones.

4.6. Influence of years of practice, age of inception, and absolute pitch

The present results do not demonstrate any clear link between frequency discrimination performance and either the age at which musical practice started, or the number of years of instrumental practice. The apparent lack of influence of the latter variable could be due to our decision to include only musicians with at least 10 years of experience. We also found no hint of a difference in pitch discrimination thresholds between the musicians who reported having absolute pitch and those who did not. However, since we did not carry out formal tests of absolute pitch in this study, it would be unwise to draw any strong conclusion on this point.

4.7. A marginal left-ear advantage in musicians

A subsidiary finding of the present study relates to the question of inter-aural differences and hemispheric dominance in pitch processing. Specifically, a left-ear advantage was observed in musicians in a condition that involved pure tones with contralateral on-frequency noise. This is partly consistent with data in the literature, which suggest a left-ear/right-hemisphere advantage in tasks involving pitch perception (Sidtis, 1980, 1981; Zatorre et al., 1992; Zatorre and Samson, 1991). The fact that a left-ear advantage was only observed in a condition involving contralateral noise occupying the same frequency region as the test tones, but not when the noise was either absent or occupied a different frequency region, is consistent with the notion that inter-hemispheric competition is required in order to evidence hemispheric dominance in the processing of auditory information (Kimura, 1964). On the other hand, it is unclear why a left-ear advantage for pitch processing was found specifically with pure tones, and not with harmonic complex tones, and why it was observed only in musicians. Clearly, the question of ear asymmetries in pitch discrimination requires further investigation. However, what we can conclude based on the present results is that, when they are observed, the differences in frequency discrimination thresholds between left and right ears are extremely subtle.

Acknowledgements

This work was supported in part by a grant from the French Ministry of Research, teaching and Education (Project # A62, Art and Cognition, ACI Cognitique 2000) to the first and third authors (then both at UMR CNRS 5020) and by NIH Grant R01 DC 05216 to A.J. Oxenham. The authors are grateful to Drs. B.J. May, R. Zatorre, and two anonymous reviewers for detailed comments on an earlier version of the manuscript, to J.G. Bernstein and C.E. Lewiston for insightful discussions, and to A. Poirier, Director of the National Music Conservatory of Paris, for his help in setting up this study. The musician and non-musician listeners are gratefully acknowledged for their willingness to participate.

References

Amitay, S., Hawkey, D.J., Moore, D.R., 2005. Auditory frequency discrimination learning is affected by stimulus variability. Percept. Psychophys. 67, 691–698.

Ari-Even Roth, D., Amir, O., Alaluf, L., Buchsenspanner, S., Kishon-Rabin, L., 2003. The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear. J. Basic Clin. Physiol. Pharmacol. 14, 137–150.

Ari-Even Roth, D., Avrahami, T., Sabo, Y., Kishon-Rabin, L., 2004. Frequency discrimination training: is there ear symmetry. J. Basic Clin. Physiol. Pharmacol 15, 15–27.

Bever, T.G., Chiarello, R.J., 1974. Cerebral dominance in musicians and non-musicians. Science 185, 537–539.

Campbell, R.A., Small, A.M., 1963. Effect of practice and feedback on frequency discrimination. J. Acoust. Soc. Am. 35, 1511–1514.

Delhommeau, K., Micheyl, C., Jouvent, R., Collet, L., 2002. Transfer of learning across durations and ears in auditory frequency discrimination. Percept. Psychophys. 64, 426–436.

Delhommeau, K., Micheyl, C., Jouvent, R., 2005. Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. J. Assoc. Res. Otolaryngol. 6, 1–9.

Demany, L., 1985. Perceptual learning in frequency discrimination. J. Acoust. Soc. Am. 78, 1118–1120.

Demany, L., Semal, C., 2002. Learning to perceive pitch differences. J. Acoust. Soc. Am. 111, 1377–1388.

Divenyi, P.L., Efron, R., Yund, E.W., 1977. Ear dominance in dichotic chords and ear superiority in frequency discrimination. J. Acoust. Soc. Am. 62, 624–632.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Grimault, N., Micheyl, C., Carlyon, R.P., Collet, L., 2002. Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. Percept. Psychophys. 64, 189–197.

Grimault, N., Micheyl, C., Carlyon, R.P., Bacon, S.P., Collet, L., 2003. Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination. Hear. Res. 184, 41–50.

Irvine, D.R., Martin, R.L., Klimkeit, E., Smith, R., 2000. Specificity of perceptual learning in a frequency discrimination task. J. Acoust. Soc. Am. 108, 2964–2968.

Johnson, P.R., 1977. Dichotically stimulated ear differences in musicians and non-musicians. Cortex 13, 385–389.

Johnsrude, I.S., Penhune, V.B., Zatorre, R.J., 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123, 155–163.

Kimura, D., 1964. Left-right differences in the perception of melodies. Q. J. Exp. Physiol. 16, 355–358.

Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y., 2001. Pitch discrimination: are professional musicians better than non-musicians? J. Basic Clin. Physiol. Pharmacol. 12 (2 Suppl), 125–143.

Lockhead, G.R., Byrd, R., 1981. Practically perfect pitch. J. Acoust. Soc. Am. 70, 387–389.

Messerli, P., Pegna, A., Sordet, N., 1995. Hemispheric dominance for melody recognition in musicians and non-musicians. Neuropsychologia 33, 395–405.

Miyazaki, K., 1989. Absolute pitch identification: effects of timbre and pitch region. Music Percept. 7, 1–14.

Moore, B.C., 1973. Frequency difference limens for short-duration tones. J. Acoust. Soc. Am. 54, 610–619.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9, 97–113.

Peretz, I., Babaï, M., 1992. The role of contour and intervals in the recognition of melody parts: evidence from cerebral asymmetries in musicians. Neuropsychologia 30, 227–292.

Peretz, I., Morais, J., 1983. Task determinants of ear differences in melody processing. Brain Cogn. 2, 313–330.

Sidtis, J.J., 1980. On the nature of the cortical function underlying right hemisphere auditory perception. Neuropsychologia 18, 321–330.

Sidtis, J.J., 1981. The complex tone test: implications for the assessment of auditory laterality effects. Neuropsychologia 19, 103–111.

Spiegel, M.F., Watson, C.S., 1984. Performance on frequency-discrimination tasks by musicians and non-musicians. J. Acoust. Soc. Am. 76, 1690–1695.

Wier, C.C., Jesteadt, W., Green, D.M., 1977. Frequency discrimination as a function of frequency and sensation level. J. Acoust. Soc. Am. 61, 178–184.

Wright, B.A., Fitzgerald, M.B., 2005. Learning and generalization of five auditory discrimination tasks as assessed by threshold changes. In: Pressnitzer, D., de Cheveigne, A., McAdams, S., Collet, L. (Eds.), Auditory Signal Processing: Physiology, Psychoacoustics, and Models. Springer, New York.

Zatorre, R.J., Samson, S., 1991. Role of the right temporal neocortex in retention of pitch in auditory short-term memory. Brain 114, 2403–2417.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849.