The Musical Ear Test, a new reliable test for measuring musical competence

Mikkel Wallentin a,b,⁎, Andreas Højlund Nielsen b, Morten Friis-Olivarius a,c, Christian Vuust d, Peter Vuust a,d

a Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Nørrebrogade, 8000 Aarhus C, Denmark
b Center for Semiotics, University of Aarhus, Denmark
c Decision Neuroscience Research Group, Copenhagen Business School, Denmark
d Royal Academy of Music, Aarhus, Denmark

A R T I C L E   I N F O

Article history:
Received 10 August 2009
Received in revised form 4 December 2009
Accepted 4 February 2010
Available online xxxx

Keywords:
Musical abilities
Musical expertise
Testing
Melody
Rhythm

A B S T R A C T

This paper reports results from three experiments using the Musical Ear Test (MET), a new test designed for measuring musical abilities in both musicians and non-musicians in an objective way with a relatively short duration (~20 min.). In the first experiment we show how the MET is capable of clearly distinguishing between a group of professional musicians and a group of non-musicians. In the second experiment we demonstrate that results from the MET are strongly correlated with measures of musical expertise obtained using an imitation test. In the third experiment we show that the MET also clearly distinguishes groups of non-musicians, amateurs and professional musicians. The test is found to have a large internal consistency (Cronbach alpha: 0.87). We further show a correlation with amount of practice within the group of professionals as well as a correlation with a forward digit span test.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Recently, the number of cognitive research studies on music has been growing, resulting in a heightened public awareness, a growing number of research communities and international conferences as well as in an increasing number of publications. A consistent finding in this literature is that music is processed differently according to level of musical expertise. This has been evidenced in behavioral studies (Bever & Chiarello, 1974; Musacchia, Strait, & Kraus, 2008; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005), with functional brain imaging techniques, such as fMRI, EEG and MEG (Gaab & Schlaug, 2003; Koelsch, Schröger, & Tervaniemi, 1999; Pantev, Roberts, Schulz, Engelien, & Ross, 2001; Vuust et al., 2005) and has been tied to organizational differences in cortical and subcortical structure (Bangert et al., 2006; Gaser & Schlaug, 2003; Hutchinson, Lee, Gaab, & Schlaug, 2003; Schlaug, Jäncke, Huang, Staiger, & Steinmetz, 1995; Schlaug, Jäncke, Huang, & Steinmetz, 1995).

In spite of this well-documented correlation between musical expertise and music processing it is striking that many, if not most, cognitive investigations of music actually leave out individually measures of musical expertise (e.g. Bangert et al., 2006; Blood & Zatorre, 2001; Chan, Ho, & Cheung, 1998; Gaser & Schlaug, 2003; Green et al., 2008; Helmbold, Rammsayer, & Altenmüller, 2005; Ho, Cheung, & Chan, 2003; Hutchinson et al., 2003; Pantev et al., 2001; Schellenberg, 2004; Schlaug, Jäncke, Huang, Staiger et al., 1995; Schlaug, Jäncke, Huang, & Steinmetz, 1995 to name but a few). The reason for this may be that an easily available, well-documented, objective, short, up-to-date test that reliably distinguishes musical expertise is lacking. In the following we will present a novel test of musical expertise, The Musical Ear Test (MET), that meets these criteria, and in three experiments show that it: I) clearly distinguishes even small groups of professional-, amateur- and non-musicians; II) correlates with the imitation test, used both at the academies of music in Denmark at entrance, mid-term and final exams, and in several studies of musical expertise (Chakravarty & Vuust, 2008; Vuust et al., 2005, 2006; Vuust, Roepstorff, Wallentin, Mouridsen, & Østergaard, 2006); and III) correlates with amount of musical practice.

1.1. Seashore’s Measure of Musical Talents

The greatest influence on the development of musical tests has been the work of Charles Seashore (e.g. see Carson, 1998). Seashore tried to form a basis for the psychological aspects of music, listed as “pitch”, “loudness”, “time”, and “timbre” on the physical properties of sound which he defined as “frequency”, “intensity”, “duration” and “wave form” (Seashore, 1942). Seashore was the first to attempt tests organized around the primitive sensory abilities that he considered the basis for musical talent (Carson, 1998). More complex structures, such as harmony, melody and rhythm were thought to be built from these elemental sensory components. Seashore’s Measure of Musical Talents (Seashore, Lewis, & Saetveit, 1960) remains the starting point for later tests, such as the Advanced Measure of Musical Audiation (Gordon, 2007 — see below), but is itself out of print (Carson, 1998).
1.2. Montreal Battery of Evaluation of Amusia

Another widely used test is the Montreal Battery of Evaluation of Amusia (MMEA) (Peretz, Champod, & Hyde, 2003). This test incorporates six subtests, five of which require the subject to make same/different judgments on two musical phrases of which the second may or may not contain a musical violation (contour, scale, interval, rhythm or metrics). The last test is a memory test in which participants judge whether a musical phrase was played in the preceding tests. The whole test takes an hour and a half to complete and as the title indicates it is primarily constructed for the study of participants with lower levels of musical abilities, such as patients with congenital amusia. A normal non-musically trained sample has a mean accuracy of 88% (corresponding to a mean score of 27/30 items on each of the subtests) and it is therefore likely to exhibit ceiling effects when used on skilled musicians.

1.3. The imitation test (the Academies of Denmark)

The imitation test has been used for around 30 years, both as a tool for identifying musical talent and for evaluating progress during education at the ‘rhythmic’ departments of the academies of music in Denmark (Improvisational styles of music such as jazz, rock, pop, reggae are by Danish tradition termed ‘rhythmic music’). Note that this does not imply that classical music is not rhythmic). In this test participants reproduce short rhythmical and melodic phrases using hand-claps and singing. This provides a short and reliable measure of musicians’ proficiency in the two fundamental aspects of music, rhythm and pitch (Palmer & Krumhansl, 1990), although the causal relationship has not been scientifically investigated. This test has no ceiling effects and is advantageously short (<10 min). However, as participants are asked to produce rhythms and melodies this test is not suitable for testing musical perception in amateur- and non-musicians. It will often yield a floor effect (see e.g. the non-musician group in Vuust et al., 2005). For experimental purposes the imitation test may also not be optimal either since it involves an element of subjective judgment of participants’ imitation.

1.4. Advanced Measure of Musical Audiation

The test that comes closest to our new test is Gordon’s “Advanced Measure of Musical Audiation” (AMMA) (Gordon, 2007). The AMMA is a same/different test that tests for melodic and rhythm deviations. Compared to the Musical Ear Test the AMMA test is primarily meant as a tool for assessing musical talent (or “aptitude”). One of the other main differences between the AMMA test and the MET is that both rhythm and melodic deviations are embedded in the same melodic context in the AMMA test, whereas the MET tests melody and rhythm separately. The AMMA approach restricts the rhythmical phrases to variations that are suitable for melody and as a consequence the rhythm part of this test doesn’t have the rhythmic variation found in contemporary styles of music (e.g. polyrhythm, integral to musical styles such as jazz and Latin). Furthermore, this approach, even though potentially more ecologically valid, introduces an attentional ambiguity in that attending to either rhythm or melody may alter the results. Curiously, when used, researchers sometimes report results from only one subtest (Schneider et al., 2002; Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005).

Another attempt at creating a test for musical abilities was done by Edwards et al. (Edwards, Challis, Hankinson, & Pirie, 2000), but this is also a rather lengthy test with several subtests and a detailed report of its practical usability has not been published. We have therefore felt the need for a new, short, reliable and easily accessible test for assessing musical competence in the normal adult population, applicable for both highly skilled musicians and for those without any musical background.

1.5. The Musical Ear Test

The newly developed Musical Ear Test consists of 104 trials on which participants judge whether two short musical phrases are identical or not (see Fig. 1). We knew that the test to be as short and simple as possible, we focused on two very fundamental aspects of music: melody (including pitch and contour) and rhythm (Krumhansl, 2000). The melodic subtest contains 52 pairs of melodic phrases, played with sampled piano sounds and the rhythm subtest contains 52 trials with rhythmical phrases, played with wood block sound. Participants respond to each trial by crossing one out of two boxes on a response sheet. Before each subtest participants are given two example trials with feedback. Half of the trials in each session (26) are “same” trials and half are “different” trials. The order of this is randomized in both sessions.

The melodies in the melody session contain 3–8 tones. Melodies have a duration of one measure and are played at 100 bpm (beats per minute). The “different” trials contain one pitch violation. For half of these (13 trials), the pitch violation also constitutes a contour violation. 25 trials contain non-diatomic tones. Of the remaining 27 trials, 20 are in the Major and 7 are in the Minor key. The order of occurrence of these features is randomized, and therefore also the difficulty.

The rhythm sequences contain 4–11 wood block beats. All sequences have a duration of one measure played at 100 bpm. The “different” trials consist of one rhythmic change. Rhythmic complexity is further varied by including triplets in 21 trials. The remaining 31 trials...
trials only contain even subdivisions of the beat. 37 trials begin on the downbeat whereas the rest begin later. The order of occurrence of these features is randomized.

No feedback is given during the test itself. The MET takes 18 min to complete. Versions of the test with English and Danish instructions are freely available from the authors and we encourage others to provide versions for use in other languages as well.

In this paper we report results from three experiments conducted to validate the MET. In the first experiment we show how the MET is capable of distinguishing between a group of professional musicians and a group of non-musicians. In the second experiment we demonstrate that results from the MET are strongly correlated with measures of musical expertise obtained using the imitation test (see above). In the third experiment we show that the MET also clearly distinguishes groups of non-musicians, amateurs and professional musicians with practically no ceiling or floor effects. We further show a correlation with amount of practice within the group of professionals as well as a correlation with a forward digit span test for the whole sample.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Forty participants took part in experiment 1, 20 (10 male; 10 female) were rhythmic musicians, all students at the Royal Academy of Music in Aarhus, 20 (9 male; 11 female) were non-musicians, 18 of which were students at University of Aarhus, one was a pedagogue, and one was a craftsman. No other demographic data were recorded.

2.1.2. Stimuli and procedure

The Musical Ear Test was conducted using a laptop computer. Stimuli were presented through headphones.

2.2. Results

A strong effect of group was found for the Musical Ear Test, both for the full test and for the melody and rhythm subtests individually (see Table 1, Fig. 2). Gender was included as a covariate in the analysis of the full model, but no effect was observed ($F(1,37) = 2.21, P = 0.14$). Performance on the two subtests was found to be correlated ($n = 40, r = 0.66, P < 0.0001$).

2.3. Interim discussion 1

In this experiment we have demonstrated the aptness of the Musical Ear Test for distinguishing expert musicians from non-musicians. The test is relatively short, requires no subjective interpretation, it can be performed by both groups of participants, making the test ideal for studies comparing expert groups with control groups. We observed no effect of gender. However, experiment 1 results cannot say anything about the relationship between the MET and musical performance abilities. Musical abilities are of course not limited to auditory discrimination, but we predict that motor capacity and auditory discrimination ability will develop together as a consequence of musical training. If so, then the MET could be considered a surrogate test for both abilities. In order to examine whether this was the case we conducted a new experiment in which participants were tested on both the MET and on the musical imitation test.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Twenty-one participants took part in this experiment. Sixteen were professional jazz/rock musicians (students at the Royal Academy of Music in Aarhus), five participants were non-musicians (see Table 2 for details on age and gender).

3.1.2. Stimuli and procedure

The imitation test consisted of 60 trials on which participants heard a short musical phrase and subsequently had to reproduce it. On 30 trials participants reproduced melodic phrases verbally and on 30 trials participants reproduced rhythmical phrases using hand-claps. Musical phrases were presented using loudspeakers connected to a laptop computer. Performance on each trial was evaluated by the experimenter as being correct (1 point) or incorrect (0 points). The imitation test has been used at the entrance exams at Danish academies of music for more than 30 years. Note, however, that it has not been formally evaluated.

Procedures for the Musical Ear Test were identical to those of experiment 1.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Non-musicians (n = 20, 11 females)</th>
<th>Professional musicians (n = 20, 10 females)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>MET — melody</td>
<td>69.7%</td>
<td>11.1</td>
</tr>
<tr>
<td>MET — rhythm</td>
<td>70.6%</td>
<td>8.0</td>
</tr>
<tr>
<td>MET — total</td>
<td>70.1%</td>
<td>7.9</td>
</tr>
</tbody>
</table>

* Model with gender as a covariate. See main text for further details.
that the MET can be used to distinguish groups of expert musicians

3.2. Results

The imitation test exhibited a strong group effect, both on the full test and on the melody and rhythm subtests (see Table 2).

The MET again exhibited a strong effect of group, for both the melody and the rhythm subtests and for the full test (see Table 2). Gender and age were included as covariates in the analysis of the full model, but effects were neither observed for age ($F(1,17) = 0.39, P > 0.5$) nor for gender ($F(1,17) = 0.39, P > 0.5$). Performance on the two subtests was again found to be correlated ($r = 0.68, P < 0.0008$).

There was a strong correlation between the imitation test and the MET ($n = 21, r = 0.89, P < 0.00001$) (Fig. 3). This effect was present both within the group of professional musicians ($n = 16, r = 0.56, P < 0.03$) and within the group of non-musicians ($n = 5, r = 0.94, P < 0.02$). There was also a performance correlation between the two melody subtests ($n = 21, r = 0.88, P < 0.0001$) and between the two rhythm subtests ($n = 21, r = 0.69, P < 0.0007$). These correlation subtest effects were also present in both professionals (melody, $n = 16, r = 0.71, P < 0.003$; rhythm, $n = 16, r = 0.5, P < 0.05$) and non-musicians, although only as a tendency in the rhythm subtest (melody, $n = 5, r = 0.95, P < 0.02$; rhythm, $n = 5, r = 0.83, P < 0.09$).

3.3. Interim discussion 2

In this experiment we replicated the finding from experiment 1 that the MET can be used to distinguish groups of expert musicians from non-musicians. We also found a strong correlation between the MET score and the score on the imitation test. The imitation test also clearly distinguished the two groups. Expert musicians were clearly better at reproducing musical phrases than non-musicians, and we further observed a strong correlation between these two tests, suggesting that auditory discrimination abilities may be used as a surrogate measure for general musical competence. This may be advantageous because the imitation test has a subjective component (performance has to be evaluated by the experimenter) and because participants with no musical training are difficult to test with this procedure. In order to further validate that the MET is linked to level of musical training it would be interesting to see if an effect of the amount of practice can be observed on the test. At least two parameters could be interesting in that respect: First, how many hours does the practitioner actually spend playing his or her instrument presently. Many studies on expertise show that practice is the main component for becoming highly skilled (e.g. Ericsson, Krampe, & Tesch-Römer, 1993; Howe, Davidson, & Sloboda, 1998; Sloboda, Davidson, Howe, & Moore, 1996). Second, at what age did he or she begin playing? An early start has been argued to be important for obtaining a very high level of musical proficiency (Schlaug, Jäncke, Huang, Staiger, et al., 1995).

A weakness in the two first experiments was that participants were not formally matched in terms of their educational level. Since all participants in experiment 1 (apart from two) were students, we would not expect that there would be a bias in our samples with regards to this parameter. But there might be effects on the MET that we are missing by only including these putatively very homogenous groups. The same thing applies to age. Although we did not see an effect of age, such an effect might actually exist if one includes a more diverse age group. Age has been found to influence performance on other working memory measures, e.g. digit span (Bopp & Verhaeghen, 2005; Dobbs & Rule, 1989). A further question that one could be interested in asking is whether the MET is capable of distinguishing a group of non-musicians from a group of amateur musicians and/or whether it can also be used to distinguish groups of amateurs and professional musicians.

With a more diverse sample we would also be able to obtain a measure of the internal consistency (such as Cronbach alpha (Cronbach, 1951)) of the MET as a test. Further we might analyze how the melodic and rhythmic variations added to the test (e.g. number of tones/beats, contour violations, triplets — see Introduction for details) contribute to the distribution of errors across trials.

Lastly, with a good measure of musical working memory, such as the MET, it becomes possible to test whether participants who have good musical ears also excel on other tests of auditory short-term memory, such as a standard digit span test. A correlation between music skills and digit span has been observed in 4–5 year-old children (Anvari, Trainer, Woodside, & Levy, 2002). Musicians have been found to develop a better verbal memory (Chan et al., 1998; Ho et al., 2003) including memory for musical lyrics, regardless of knowing the
melody (Kilgour, Jacobson, & Cuddy, 2000). One study also reported borderline significant effects of musical training on verbal working memory (Franklin et al., 2008). Jacobson, Cuddy, and Kilgour (2003) found a positive correlation between years of musical training and short-term retention of syllables and suggested that music training strengthens auditory temporal-order processing abilities. Our last study was designed to investigate all of these parameters while at the same time aiming at performing another replication of the finding that the MET is a viable test for measuring musical competence across groups of different proficiencies.

4. Experiment 3

4.1. Methods

4.1.1. Participants

Sixty participants took part in this experiment, 18 professional musicians, 21 amateur musicians and 21 non-musicians (see Table 3).

Professional musicians were defined as participants who either made a living by playing music and/or who were enrolled in or educated from one of the musical academies. Thirteen professionals considered themselves jazz/rock musicians, five considered themselves primarily classical musicians. Amateur musicians were considered as those who had played an instrument for at least two years and who for the last year had spent on average at least one hour per week practicing/play. Age of onset and the average amount of weekly playing over the last year were used covariates for analyses related to effects of practice. Participants who had no current musical activities were considered non-musicians, but we tried to avoid including participants who had a long prior history of playing, however to be consistent participants were not excluded if such a past was discovered during the experiment (this happened only once). All participants spoke Danish as their first language.

4.1.2. Stimuli

The forward digit span test was conducted as follows. Starting with four digits, participants repeated a sequence of numbers read aloud to them immediately after presentation. The numbers were read by the experimenter with an interdigit interval of approximately one second. At each level (i.e. number of digits) participants were given four trials. In order to proceed to the next level participants were required to repeat the correct sequence in at least three out of four trials. Upon making two errors, participants received the digit span according to the previous level (on which maximally one error had occurred).

The Musical Ear Test was identical to that of experiments 1 and 2.

4.1.3. Analysis

Group differences between non-musicians, amateurs and professional musicians on the MET and the digit span test were analyzed using a Non-parametric test. Chi-square-score. For the melody subtest alone it was 0.94. This indicates a large internal consistency for the test as a whole as well as for the individual subtests.

To test the effect of the different kinds of variations added to the trials (see Introduction) we computed the average accuracy across participants for each trial. We then used this measure as a dependent variable in two ANOVA analyses, one for the melody and one for the rhythm session.

The ANOVA analysis of the melody session trials revealed a marginally significant effect of number of tones (F(1,47) = 3.77; P < 0.06), a significant effect of violation/violation type (F(2,47) = 30.01; P < 0.001) and a significant effect of presence of non-diatonic tones (F(1,47) = 11.88; P < 0.002). No interactions were observed. Post hoc comparisons using Tukey’s honestly significant difference (HSD) test revealed that trials including only pitch violation were significantly more difficult (mean accuracy: 0.57; s.e.: 0.03) than both “same” trials (mean accuracy: 0.81; s.e.: 0.02) and trials with both pitch and contour violations (mean accuracy: 0.89; s.e.: 0.03). No significant difference was observed between “same” and pitch/contour violation.

The ANOVA analysis of the rhythm session trials found that the main effect of number of beats was non-significant (F(1,48) = 0.51; P = 0.4), whereas a significant effect of the presence of triplets was found (F(1,48) = 8.13; P < 0.01) and a significant interaction between number of beats and the presence of triplets in the trials was also found (F(1,48) = 5.41; P < 0.03).

4.2. Forward digit span

All participants’ digit span scores fell between 4 and 8. Due to this lack of continuity in the data we analyzed group effects using a non-parametric ANOVA (Kruskal–Wallis). A significant effect of group was found (Chi-square (2,57) = 6.59, P < 0.05). Post hoc comparisons using Tukey’s honestly significant difference (HSD) test indicated that non-musicians had a significantly lower digit span than professional musicians, whereas there were no differences between non-musicians and amateurs and between amateurs and professionals (see Table 3, Fig. 4).

4.2.3. Musical Ear Test

Again, we found a very strong effect of group on the Musical Ear Test (see Table 3, Fig. 4). Planned post hoc tests revealed a significant difference between all groups (non-musicians vs. amateurs (t(40) = 5.39, P < .0001; non-musicians vs. professionals (t(37) = 10.11, P < .0001;

Table 3

Experiment 3: demographics and results.

<table>
<thead>
<tr>
<th></th>
<th>Non-musicians (n=21, 9 females)</th>
<th>Amateur musicians (n=21, 7 females)</th>
<th>Professional musicians (n=18, 5 females)</th>
<th>F (df)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>M=30.2, SD=11.1</td>
<td>M=28.0, SD=11.2</td>
<td>M=30.4, SD=6.8</td>
<td>0.36 (2.57)</td>
<td>&lt;0.68</td>
</tr>
<tr>
<td>Education (years)</td>
<td>M=17.3, SD=2.9</td>
<td>M=16.7, SD=2.8</td>
<td>M=18.3, SD=2.7</td>
<td>1.63 (2.57)</td>
<td>&lt;0.20</td>
</tr>
<tr>
<td>Age onset playing</td>
<td>M=12.8, SD=5.3</td>
<td>M=9.3, SD=3.8</td>
<td>M=9.3, SD=3.8</td>
<td>5.30 (1.37)</td>
<td>&lt;0.03</td>
</tr>
<tr>
<td>Weekly playing (hours)</td>
<td>M=4.7, SD=4.7</td>
<td>M=25.53, SD=1.37</td>
<td>M=25.53, SD=1.37</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Forward digit span</td>
<td>M=5.67, SD=1.06</td>
<td>M=6.24, SD=0.94</td>
<td>M=6.39, SD=0.6</td>
<td>6.59 (2.57)</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>MET – melody</td>
<td>M=66.7%, SD=8.4</td>
<td>M=78.6%, SD=7.5</td>
<td>M=89.1%, SD=5.6</td>
<td>45.65 (2.57)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>MET – rhythm</td>
<td>M=69.0%, SD=8.7</td>
<td>M=77.4%, SD=7.1</td>
<td>M=82.3%, SD=7.5</td>
<td>14.67 (2.57)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>MET – total</td>
<td>M=67.8%, SD=6.1</td>
<td>M=78.0%, SD=6.0</td>
<td>M=85.7%, SD=4.6</td>
<td>48.15 (2.54)</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>

* Non-parametric test. Chi-square-score.

b Model with gender, age and education as covariates. See main text for further details.
amateurs vs. professionals \((t(37)=4.42, P<0.0001;\text{ see Fig. 4})\). A small, but significant effect of age was also observed \((F(1,54)=4.83, P<0.05)\).

There were no effects of gender or level of education \((P>0.05\) for both covariates). Performance on the two subtests was again found to be correlated \((n=60, r=0.52, P<0.0001)\).

The amateur and professional musicians differed both in age of onset and weekly hours of playing \((Table 3)\). Scores across these two groups on the Musical Ear Test were strongly correlated with weekly hours of practice and playing \((n=39, r=0.55, P<0.005)\) \((Fig. 5)\). Within the group of professionals alone we also found a correlation between scores on the Musical Ear Test and amount of weekly playing \((n=18, r=0.52, P<0.05)\), whereas this was not the case for amateurs \((n=21, r=0.04, P>0.85)\).

Age of onset was found not to be significantly correlated with the score on the Musical Ear Test for the whole group of musically trained participants \((n=39, r=0.24, P>0.2)\). A marginally significant correlation was found within the group of professional musicians \((n=18, r=-0.46, P<0.06)\), indicating that an early start also has an effect on proficiency in the high end of the MET scale. Amount of practice and age of onset were not mutually correlated within the professionals \((n=18, r=-0.23 P>0.35)\), suggesting that these may indeed be independent effects.

Finally, we found a positive correlation between digit span and full score on the Musical Ear Test \((n=60, r=0.40, P<0.002)\). This correlation was also present across both the melody \((n=60, r=0.29, P<0.03)\) and the rhythm \((n=60, r=0.42, P<0.001)\) subtests alone. These effects were not present within any of the groups individually \((P>0.05,\text{ for all tests})\).

4.3. Interim discussion 3

Again, we found that the MET is capable of distinguishing groups of non-musicians from groups of professional musicians. We further showed that the test is capable of clearly distinguishing non-musicians from a group of amateur musicians and of distinguishing the amateurs from the professionals. Neither gender nor level of education was found to affect the results whereas we found a small effect of age. Working memory has been found to decline slightly for ages above 60 years \((Bopp & Verhaeghen, 2005; Dobbs & Rule, 1989)\), and this effect seems also to be applicable to the MET.

4.3.1. Effects of practice

We found that performance on the MET was significantly correlated with the amount of current weekly practice/playing. This effect was present also in the group of professionals alone, but not in the amateurs, suggesting that up until a certain level of proficiency one may rely on either general intelligence \((Lynn, Wilson, & Gault, 1989; Madison, Forsman, Blom, Karabanov, & Ullén, 2009; Ramm-sayer & Brandler, 2007)\), talent or past effort, but to go above that requires a certain amount of ongoing practice \((e.g. Ericsson et al., 1993; Howe et al., 1998; Sloboda et al., 1996)\). It has been found that high achievers in music do not differ from the less successful ones with respect to how fast they acquire new skills \((Sloboda et al., 1996)\). This limits the role for musical talent and puts an emphasis on the role of extensive practice. Previous studies have shown that practice is accompanied by plastic changes to the brain both in white \((Bengtsson...)}
et al., 2005; Schlaug, Jäncke, Huang, Staiger, et al., 1995) and in gray (Gaser & Schlaug, 2003) matter. With efficient testing procedures for musical abilities, such as the MET, we predict that such effects can be studied in much greater detail. A tendency for early onset of playing to play a role in MET score levels further indicated that this may also be important for acquiring very high levels of proficiency.

4.3.2. Transfer effects
The idea that music listening affects other cognitive abilities has a long history. The most famous case is the so-called “Mozart effect” discovered by Rauscher et al. (Rauscher, Shaw, & Ky, 1993) who found that even brief exposure to music can have measurable impact on a variety of cognitive and neurological processes (see also Ivanov & Geake, 2003; Jausovec, Jausovec, & Gerlic, 2006; Zhu et al., 2008). The most commonly found explanation is that the Mozart effect is due to the effects of music on arousal and mood (Chabris, 1999; Nantais & Schellenberg, 1999; Steele, Ball, & Runk, 1997; Steele, Bass, & Crook, 1999; Thompson, Schellenberg, & Husain, 2001; Vuust & Kringlebch, 2010). That this effect should be specific to the music of Mozart or even to music per se is, however, highly disputed. In a study on 10- and 11-year-olds, Schellenberg and Hallam (2005) found an effect of listening to pop music but no effect of listening to Mozart. This study indicates that positive benefits of music listening on cognitive abilities are dependent on how much listeners enjoy the music. This is corroborated by Nantais and colleagues who showed that performance on an IQ test was a function of listener’s ‘preference’, no matter whether this preference was for Mozart, Schubert or for listening to a story (Nantais & Schellenberg, 1999).

Musical transfer effect of a more limited scope, however, may be more consistent. In this experiment we observed a weak correlation between score on the MET and a forward digit span measure. This correlation was more consistent. In this experiment we observed a weak correlation between score on the MET and a forward digit span measure. (see also Ivanov & Geake, 2003; Jausovec, Jausovec, & Gerlic, 2006; Zhu et al., 2008). The MET showed no ceiling effects even with highly proficient professional musicians. No participants were 100% accurate on the full Musical Ear Test. Only two participants out of all three experiments scored 100% on one of the subtests (one reaching ceiling on the melody subtest, the other on the rhythm subtest).

5.3. Floor effects
In order to score significantly above chance ($P<0.05$) participants must have more than 60 out of 104 trials correct on the full MET. Only two participants out of all three experiments ($n=120$) fell below this threshold. In order to be above chance level on the subtests, participants must score 32 out of 52 trials correctly. 19 participants fell below this threshold on one subtest (11 on the melody and 8 on the rhythm subtest). One should therefore be cautious when interpreting the subtest results from non-musicians.

5.4. Performance differences on subtests
Performance on the MET subtests (melody and rhythm) was strongly correlated through all experiments. We have yet to determine whether the two subtests may be sensible to different aspects of musical expertise, e.g. we predict that musicians with absolute pitch will have a high score on the melody subtest, but not necessarily have an equally high score on the rhythm subtest. Indeed, we are aware of one such case in our study. This participant had an accuracy of 98% on the melody test, but scored just above chance on the rhythm test (61.5%). To our knowledge, this participant was the only one with absolute pitch in our study.

Playing certain instruments may also cause musicians to excel on one subtest, e.g. drummers might score higher on the rhythm subtest (Rauscher & Hinton, 2003). Finally, we might expect jazz/rock musicians in general to outperform classical musicians on the rhythm subtest, whereas classical musicians may be better on the melody subtest. In experiment 3 we included five classical musicians. We did not detect any difference between jazz/rock and classical musicians ($P>0.05$, data not shown), but a larger study is needed to determine if this null-finding is replicable.

5.5. Working memory issues
One possible limitation of this test is that it loads heavily on working memory. For more extensive testing of subjects’ musical abilities it might be useful to include extra tests that require less maintenance or manipulation of musical information in working memory. But again, the wish for a relatively brief and simple test, easily evaluated even by people with no musical background, weighs in favor of this type of comparison paradigm. Further, it must be added that music perception and production are rarely independent of working memory, since both melody and rhythm function in relation...
to pre-specified reference frames (harness and meter) that have to be maintained and/or manipulated within working memory in order for music to be meaningful (e.g. see Vuust et al., 2005).

6. Conclusion

With a test lasting only about 20 min we obtain a score with hardly any overlap between distributions for non-musicians and professional musicians. This relatively short test has a large internal consistency (Cronbach alpha: 0.87; melody subtest: 0.96; rhythm subjects: 0.94). It furthermore produced results that were comparable with level of ability for reproducing musical phrases and performance was also correlated with individual amounts of practice/playing. Both of these results demonstrate a strong connection between MET score and real musical abilities. The test further has the advantages of being applicable to both musicians and non-musicians exhibiting only a very small floor effect and no ceiling effect, even in highly proficient musicians. The results seem not to be affected by other demographic variables, such as gender or level of education while a small effect of age has to be taken into account. This study also found a small effect of playing an instrument on a secondary verbal short-term measure, forward digit span. This replicates earlier findings of a weak transfer effect of musical training onto other auditory domains. All in all the MET has proved to be a potentially very valuable tool for researchers investigating music cognition and its neural correlates.

Acknowledgements

This study was supported by The Carlsberg Foundation, The Ulla and Folmer Mogensen Foundation and The Danish National Research Foundation grant to CFN.

References


