Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Tales of Talent: Rapid Learning of Acoustic Instrument Recognition

Lisa Aufegger,[*1] Oliver Vitouch[*2]

[*]*Dept. of Psychology, University of Klagenfurt, Austria*

[1]lisa.aufegger@rcm.ac.uk, [2]oliver.vitouch@aau.at

ABSTRACT

Also in the 21st century, the role of innate talents in music remains a matter of fundamental debate. Within the framework of the "rapid learning paradigm", the aim of this study was to find out whether it is possible to simply and quickly teach non-musicians musical skills in the perceptual realm, specifically the recognition of instruments' timbres. Within a week, 34 subjects had three feedback-driven computer-based training sessions, where they were asked to discriminate between 10 brass and woodwind instruments. In the pre- and a post-test, subjects had to recognize the main instrument from an orchestral piece. Results shown that non-musicians did not fully reach expert level (benchmarked by brass or woodwind instrument students) after this short period, but performed well at par with semi-experts (piano students). Our findings demonstrate that acoustic instrument recognition is well-trainable "for (almost) everybody" using the simplest of means, and does not seem to depend on rare individual abilities.

I. INTRODUCTION

The existence of musical talent has long been a matter of vigorous debate. "Talent", in this context, usually denotes abilities that are individually specific, genetically based (or "innate"), and rare (Howe, Davidson, & Sloboda, 1998). Recently, Oechslin, Läge, and Vitouch (2012) defined a "rapid learning paradigm" for the controlled empirical demonstration that a certain task is well-accessible for learning "by anybody". If a defined musical listening task, which initially cannot be adequately performed, can be easily and rapidly trained, resulting in expert-like performance after the training, then it is plausible to argue that learning plays a central role for this task, and that the task is broadly accessible, not restricted to a talented minority. If this can be shown for a multitude of tasks, it would help to empirically deconstruct a pervasive talent account of music activity, which does at least prevail in "folk psychology".

Theoretically guided speculations about what competencies the average listener can achieve through simple, short-term training may lead to quite heterogeneous conclusions. On the one hand, basic features of pitch comparisons, contour and *gestalt* perception or perception of consonance (e.g., octave similarity) seem to be universally shared by humans – an evolutionary commonality of *Homo sapiens* (cf. Vitouch & Ladinig, 2009). Congenital amusia, or "tone-deafness", is a minority problem: It is assumed to be a genetically-based cortical deficit of fine-grained pitch perception, with a frequency of about 4 % in the general population (Drayna, Manichaikul, de Lange, Snieder, & Spector, 2001; Hyde & Peretz, 2004; Peretz & Hyde, 2003).

However, the question about genetic bases of musicality, and about decisive individual differences in this respect, can strongly depend on the sub-domain in focus. While adequate relative pitch seems to be a well-trainable property (with few genetically based exceptions), the ability of absolute pitch seems to be largely unattainable by training in post-critical ages. Many authors favor an early-learning account of absolute pitch (Russo, Windell, & Cuddy, 2003; Takeuchi & Hulse, 1993; Vitouch, 2003; Ward, 1999) with a sensitive period for acquisition up to age 6. Also the characteristics of the first language acquired seem to play an eminent role as soon as tone languages (e.g., Mandarin) are involved (Deutsch, Henthorn, Marvin, & Xu, 2006; Deutsch, Dooley, Henthorn, & Head, 2009). Additionally, there may indeed be a genetic factor involved here, which remains to be discovered (Levitin & Rogers, 2005a, 2005b; Zatorre, 2003). Still, rudimentary (or "latent") absolute pitch may be much more frequent than is traditionally assumed (Gussmack, Vitouch, & Gula, 2006; Schellenberg & Trehub, 2003; Vitouch & Gaugusch, 2000).

Historically, talent has long been framed as a bundle of "inborn faculties and dispositions" (Galton, 1869), or a genetic product linked to a high general intelligence (Terman, 1925). It took decades before researchers started to consider the influence of additional factors, such as motivation, creativity (Renzulli, 1978), or social environments (Mönks, 1992). As a result, more than one factor had to be considered now in order to better comprehend the intellectual backstage of a great performance. This was also attempted by DeGroot (1978) and Chase and Simon (1973), who have shown that experts have indeed better memory skills, but these are not caused by any congenital ability but rather by practice and work (Lehmann, Sloboda, & Woody 2007), built on the 10-years rule of intense training (Ericsson, Krampe, & Tesch-Römer, 1993). This approach of attaining expertise does not assume a high IQ or rare special abilities, but instead posits the possibility of achievement as available for almost everyone through the quality of practice and motivation by means of useful strategies and their adaptability to certain exercises.

Based on these results, recent studies attempted to focus not only on this learning paradigm, but also on whether individuals without any experiences in a particular field are able to attain these skills by means of simple learning strategies. Smith, Kemler Nelson, Grohskopf, and Appleton (1994) designed training methods, used by non-musicians with great ease to recognize different kinds of intervals. Oechslin et al. (2012) could demonstrate that more than 50 percent of their subjects quickly achieved expert level, operationalized via mental maps based on non-parametric multidimensional scaling (NMDS), in a tonal similarity task with tone-chord ratings. Levitin (1994) and Levitin and Cook (1996) presented evidence for very good performance of non-musicians in terms of absolute tonality (vocal production method) and absolute tempo. Bigand (1990) found similarities between musicians and non-musicians in the abstraction of forms based on melodic variations over a set of chords.

Even Wolpert's (1990) claim of different memorization strategies for melodies could be partly refuted by Radvansky, Fleming and Simmons (1995), who revealed that timbre changes in melodies affected neither non-musicians nor musicians. Racette and Peretz (2007) demonstrated that both musicians and non-musicians seem to struggle when learning a new pop song. Based on findings like these, it has been assumed that musical skills are as basic as language (Peretz, 2006), so that everyone could be labeled as an "experienced listener" (Bigand & Poulin-Charronnat, 2006).

Although many findings exist along these lines, little research has been done so far regarding the trainability of timbre recognition with musical instruments, and of acoustic recognition of musical instruments in general. This may be due to the complexity and characteristics of timbre's dimensions (e.g., spectral centroid, Grey, 1977; attack synchronicity, spectral flux, overtone synchronicity, fluctuation strength, McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995), as well as its definition. However, some studies have attempted to develop tests for the recognition of musical instruments. Saldana and Corso (1964) emphasized the importance of the transitional information between notes, and Kendall (1986) focused on the context of musical passages so as to be able to percept musical quality and recognize instruments. McAdams et al. (1995) found that musicians and non-musicians judged musical sounds in a very similar way, being less different in their perceptions than commonly believed (McAdams et al., 1995).

Our aim was to investigate to what extent non-musicians are able to reliably discriminate different timbres within a family (wind instruments) after a brief period of feedback-based training. The post-training assessment required a transfer of learning: Instruments were presented with solo excerpts in the training phase, and as main/solo instruments with orchestral backgrounds in the post-test. The training method relies on the "rapid learning paradigm" (Oechslin et al., 2012), a feedback-based approach with assumed conceptual links to perceptual learning (e.g., Goldstone, 1998). Drawing on the notion that subjects are not consciously aware of how they are learning, implicitly establishing new distinct categories, simple feedback (correct / wrong) is used as a tool to reinforce the learning progress. The results after training were compared with ad-hoc results (no training) from different groups of music students (wind instrument players and others).

II. PARTICIPANTS

The study included a total number of $N = 53$ participants: $n = 34$ non-musicians ($M = 31$ years, $SD = 13$; 71% female), $n = 9$ brass and woodwind students ($M = 34$ years, $SD = 6$; 66% female), and $n = 10$ piano students ($M = 31$ years, $SD = 3$; 60% female). Non-musicians had no episode of music performance education in their biographies that had lasted longer than three years. At primary school, 61% of the sample had played an instrument, but no longer than 1.4 ($SD = 1.2$) years. This group was compared to brass and woodwind instrument students (experts) and to piano students (semi-experts) from the state-funded Music Conservatory in Klagenfurt, Austria. Non-musicians and piano players were compared based on the assumption that piano players have comparatively little direct experience with brass or woodwind instruments (focus on solo performance, usually no orchestra or chamber music involvement, relatively rare accompaniments of wind players as compared to string players or singers).

III. MATERIALS AND METHOD

A. General Design

In order to provide a training unit free from defects, a pre- and a posttest as well as three feedback-driven training units were created with the aid of the software „hot-potato 6" which can be found at http://hotpot.uvic.ca/ (July 17th, 2009). Length, scope and feedback rate of each training pass was determined based on a pilot study. In the pre- and posttest, subjects were presented with musical excerpts of 10 woodwind and brass instruments from orchestral recordings. The woodwind instruments contained a clarinet, a recorder, flute, saxophone, bassoon and an oboe, while the brass instruments included tuba, trombone, trumpet and a French horn. Each audio sample for the pre- and posttest lasted 20 ± 5 s, using excerpts from various classical CDs. After subjects had listened to the orchestral excerpts, they were asked to make their choice without receiving any feedback. In order to recall all the instruments in the set, subjects were handed out an illustrated paper with all relevant instruments for orientation.

After the pretest, subjects had their first training session, which was separated into two parts, starting with an explanation by means of a PowerPoint presentation of the instruments' characteristics (see sub-section B. below) and followed by the PC-based training unit.

The 80 audio samples from the training unit contained five standardized and three specific tunes per instrument, and were recorded via the ZOOM H4N-Recorder. For these audio samples, musicians from a secondary school, from the Klagenfurt Conservatory and from the Vienna Youth Philharmonic Orchestra were recruited in order to provide a high quality of musical sound.

Each sample lasted 10 ± 5 s, with the samples being duly parallelized in terms of their dynamics, volume, sound, and quality. The order of the audio samples and their response options were randomly generated. Each sample could be played as often as required. When clicking on an instrument, participants received an "x" for incorrect responses, being asked to try again until the correct answer was found, and a "☺" for correct responses (see Figure 1).

As a criterion to be given a pass on each training session, subjects were required to correctly identify 70% of the audio samples. If this was not achieved, they had to repeat the training block until they reached this percentage, or until a period of 90 minutes was over. The third and last training session was followed by the post-test.

B. Summary of the Training Units

All units were held every second day, starting with the first lesson, in which participants were provided with simple information (e.g., theory of timbre, design of the instrument) about the brass wind family through a PowerPoint presentation as well as with instructions by the researcher [L.A.]. During the

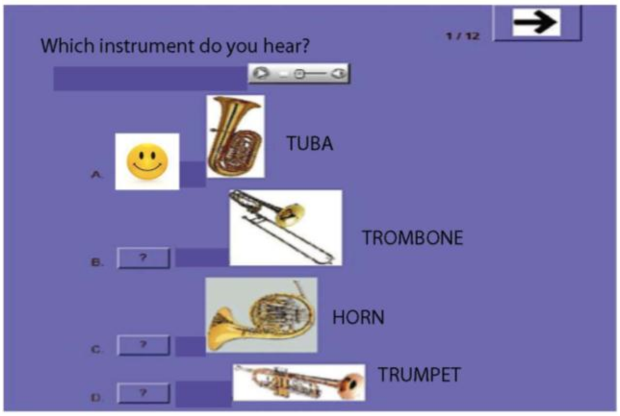

Figure 1. Screenshot of the training unit (brass instruments)

session, participants had to listen to 32 audio samples (resulting from 5 standardized and 3 specific tunes per instrument) and to correctly identify at least 23 samples in order to pass and get on to the second learning unit.

The second unit contained the same sets of conditions as the first, with the only exception of using woodwind instruments. After the instructional introduction, 48 audio samples were presented with the requirement to identify at least 34 samples.

The last unit presented both brass and woodwind audio samples and was only presented once in order to ensure enough time and concentration for the post-test (entailing transfer of learning to orchestral excerpts), which was performed directly afterwards.

IV. RESULTS

The non-musicians showed a significant ($p < .001$, $t = -10.5$, $df = 33$) improvement, recognizing 3.4 (*SD* = 1.6) instruments in the pre-test and 6.8 (*SD* = 1.8) instruments in the posttest (effect size $d = 1.86$). A detailed comparison between the brass and the woodwind instruments showed similar results for both sub-families (brass: $p < .001$, $t = -7.3$, $df = 33$; woodwind: $p < .001$, $t = -5.3$, $df = 33$). Subjects identified 2.0 (*SD* = 1.0) vs. 3.2 (*SD* = 1.0) brass instruments and 1.4 (*SD* = 1.1) vs. 3.5 (*SD* = 1.415) woodwind instruments in the pre- vs. post-test.

An analysis of variance between the groups “non-musicians”, “experts”, and “semi-experts” indicated that non-musicians could not reach the expert level (with experts answering ad-hoc, without receiving any experimental training) of 9.8 (*SD* = 0.1) hits, but achieved semi-expert level, even descriptively surpassing the performance level of piano students with 5.8 (*SD* = 1.6) hits (see Figure 2). A closer look at the latter difference shows that non-musicians recognized 2.0 (*SD* = 0.1) and pianists 3.6 (*SD* = 1.5) woodwind instruments on average, whereas non-musicians could identify 3.2 (*SD* = 1.0) brass instruments as compared to only 2.0 (*SD* = 0.8) for the pianists.

An independent *t*-test showed that woodwind recognition was statistically equal ($t = -0.38$, $df = 42$, $p = .702$), whereas non-musicians had a significantly higher hit rate with brass instruments ($t = 3.3$, $df = 42$, $p = .002$). Individual experiences of playing an instrument at primary school in the non-musicians had no significant influence on the post-test outcomes ($t = -0.88$, $p = .417$). Non-musicians with no music lessons even had a slightly better descriptive rate in the post-test than non-musicians with experience.

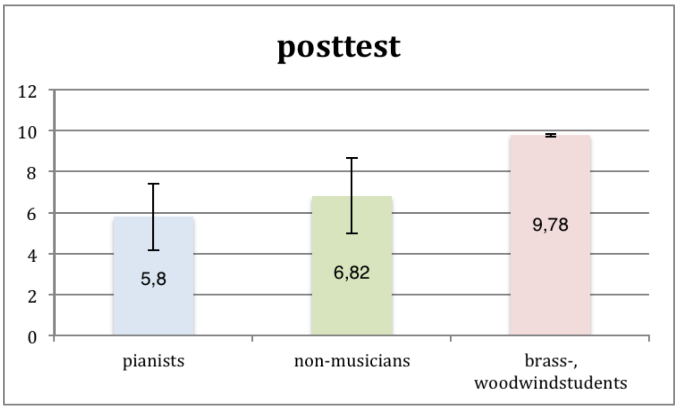

Figure 2. No. of instruments identified by pianists, non-musicians (post-test), and brass / woodwind students [bars indicate standard errors]

V. CONCLUSION

The talent versus expertise debate is a broad and long-lasting issue, holding in itself different theories about musical achievement. Whereas the former is focused on the notion of genetic potential – to an extent exclusive – the latter approaches the issue from the perspective of a strong relationship between performance and practice, with high performance attainable by almost everyone. Results from our study demonstrate that non-musicians improved markedly in their recognition performance between pre- and post-test for both woodwind and brass instruments. While they did not attain “expert level” in this domain, they performed surprisingly well in comparison to the “semi-experts” (pianists), based on a brief and simple training method.

Given the adequate methodology, and using a feedback-based approach, non-musicians are, as demonstrated, able to perceptually excel, and to even slightly surpass piano students. These results should also bend thoughts on what and how pianists are taught as well as on their interplay with other instrumental groups. Since pianists are mostly acting as soloists or as accompanying instrumentalists for certain groups of instrument families, it would seem important to amplify their range of experience, and to implement this in their curriculum.

In the context of perceptual learning, there may indeed be a broad and general basis for quickly acquiring and improving fundamental competencies of musical perception. Furthermore, a quick follow-up study showed that the subject with the best test results had the same or, in some instances, a better rate 4 to 8 weeks after the initial training sessions. Unfortunately not all of the subjects were available for this informal follow-up, meaning that only some, including the best of the group, could be tested. In addition, the same material was used in the follow-up test (familiarity effects), so we have no firm evidence as to whether these results stem from longer-term transfer of knowledge or whether this was merely due to a memory effect of their performance.

Should further studies find out that knowledge can be sustained and transferred, and that non-musicians are able to use these skills in different context, it would also be interesting to see what exactly and how much they are able to attain – how far they can go (*testing-the-limits,* Kliegl & Baltes, 1987). We do know about trainability regarding interval recognition

(Smith et al., 1994) and tonal similarity ratings (Oechslin et al., 2012), but what about rhythm and other domains? If strong effects of “quick’n’easy” training can be demonstrated for a broad range of listening tasks (and some non-instrument-bound production tasks), then one could well ask why music academies are still using unstandardized admission tests using such tasks, assuming diagnostic reliability of these tests for innate “musicality”, and predictive validity for success in music.

In conclusion, a simple feedback-driven training regimen led to surprisingly strong improvements in timbre recognition within a short period of time. This speaks for the general trainability of this task, with good performance rather being an (easily) acquired skill than a mysterious innate ability. Further research should test the “rapid learning paradigm” in other domains of music, and should provide follow-up insights into the longevity and the limits of these effects.

REFERENCES

Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology, 28,* 121-125.

Bigand, E. (1990). Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music, 18,* 45-60.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100,* 100-130.

Chase, W. G., & Simon, H. A. (1973). The mind’s eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

De Groot, A. (1978). *Thought and choice in chess.* The Hague, The Netherlands: Mouton. [Originally published in 1946.]

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

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

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science, 291,* 1969-1972.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acqusition of expert performance. *Psychological Review, 100,* 363-406.

Galton, F. (1869). *Hereditary genius.* London: Macmillan.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585-612.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61,* 1270-1277.

Gussmack, M. B., Vitouch, O., & Gula, B. (2006). Latent absolute pitch: An ordinary ability? In M. Baroni, A. R. Addessi, R. Caterina & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception & Cognition* (pp. 1408-1412; CD-ROM). Bologna, Italy: Bononia University Press.

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: reality or myth? *Behavioral and Brain Sciences, 21,* 399-442.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15,* 356-360.

Kendall, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception, 4,* 185-214.

Kliegl, R., & Baltes, P. B. (1987). Theory-guided analysis of mechanisms of development and aging through testing-the-limits and research on expertise. In C.Schooler and K. W. Schaie (eds.), *Cognitive functioning and social structure over the life course* (pp. 95-119). Norwood, NJ: Ablex.

Lehmann, A. C., Sloboda, J. A., & Woody, R. H. (2007). *Psychology for musicians: Understanding and acquiring the skills.* New York: Oxford University Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56,* 414-423.

Levitin, D. J., & Cook, P. R. (1996). Absolute memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics, 58,* 927-935.

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

Levitin, D. J., & Rogers, S. E. (2005b). Erratum: Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences, 9,* 45.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58,* 177-192.

Mönks, F. J. (1992). Ein interaktionales Modell der Hochbegabung. In E. A. Hany & H. Nickel (Eds.), *Begabung und Hochbegabung* (pp. 17-22). Bern, Switzerland: Huber.

Oechslin, M. S., Läge, D., & Vitouch, O. (2012). Training of tonal similarity ratings in non-musicians: A “rapid learning” approach. *Frontiers in Psychology, 3:*142 [doi: 10.3389/fpsyg.2012.00142]

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition, 100,* 1-32.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7,* 362-367.

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory and Cognition, 35,* 242-253.

Radvansky, G. A., Fleming, K. J., & Simmons, J. A. (1995). Timbre reliance in nonmusicians’ and musicians’ memory for melodies. *Music Perception, 13,* 127-140.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60,* 180-184.

Russo, F. A., Windell, D. L., & Cuddy, L. L. (2003). Learning the “special note”: Evidence for a critical period for absolute pitch acquisition. *Music Perception, 21,* 119-127.

Saldana, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America, 36,* 2021-2026.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14,* 262-266.

Smith, J. D., Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition, 52,* 23-54.

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

Terman, L. M. (1954). The discovery and encouragement of exceptional talent. *American Psychologist, 9,* 221-230.

Vitouch, O. (2003). Absolutist models of absolute pitch are absolutely misleading. *Music Perception, 21,* 111-117.

Vitouch, O., & Gaugusch, A. (2000). Absolute recognition of musical keys in non-absolute-pitch-possessors. In C. Woods, G. Luck, R. Brochard, F. Seddon & J. A. Sloboda (Eds.), *Proceedings of the 6th International Conference on Music Perception and Cognition* (CD-ROM). Keele, UK: Keele University Dept. of Psychology.

Vitouch, O., & Ladinig, O. (Eds.) (2009). Music and evolution. *Musicae Scientiae, Special Issue 2009-2010.*

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

Wolpert, R. S. (1990). Recognition of melody, harmonic accompaniment, and instrumentation: Musicians vs. nonmusicians. *Music Perception, 8,* 95-106.

Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience, 6,* 692-695.