Available online at www.sciencedirect.com
ScienceDirect
Neuroscience Letters 405 (2006) 164–167
Neuroscience Letters
www.elsevier.com/locate/neulet

Superior voice timbre processing in musicians

Jean-Pierre Chartrand [a,*], Pascal Belin [a,b]

[a] *BRAMS, Département de Psychologie, Université de Montréal, C.P. 6128, succursale Centre-Ville Montréal, Qué., Canada, H3C 3J7*
[b] *Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK*

Received 15 December 2005; received in revised form 10 April 2006; accepted 26 June 2006

Abstract

After several years of exposure to musical instrument practice, musicians acquire a great expertise in processing auditory features like tonal pitch or timbre. Here we compared the performance of musicians and non-musicians in two timbre discrimination tasks: one using instrumental timbres, the other using voices. Both accuracy (*d*-prime) and reaction time measures were obtained. The results indicate that the musicians performed better than the non-musicians at both tasks. The musicians also took more time to respond at both tasks. One interpretation of this result is that the expertise musicians acquired with instrumental timbres during their training transferred to timbres of voice. The musician participants may also have used different cognitive strategies during the experiment. Higher response times found in musicians can be explained by a longer verbal-auditory memory and the use of a strategy to further process auditory features.

Keywords: Voice; Musicians; Expertise; Musical training

In auditory perception, one of the most studied forms of expertise has been that of musicians. Practicing a musical instrument usually starts during childhood or adolescence and after several years of exposure these people acquire great expertise with processing of auditory features like tonal pitch or timbre [13,10]. Musical training requires and involves different kinds of auditory processes such as recognition and discrimination of instrument tones, and musicians improve all those processes over time. Musicians' expertise with processing of timbre of musical instruments has also been demonstrated [13].

Electrophysiological data show that musical training is accompanied by specific cortical modifications. The P2 and N1c (sub-component of N1 occuring at electrode T8) components of the auditory evoked potential in response to music sounds are found to be enhanced in musicians [14]. The neuromagnetic N1m component is likewise enhanced in musicians [11], especially for the musical timbres of their own instrument [12]. Moreover, when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones, the P2 component is enhanced bilaterally and the N1c is enhanced in the right hemisphere [3]. The P2 is a particularly sensitive indicator of neural plasticity since it can be enhanced early in childhood, for example when 4- to 5-year-old children are exposed to musical training [16].

Both behavioral and electrophysiological data demonstrate that changes occur in musicians and that these changes are closely associated with the processing of musical features. But to what extent can musical expertise with musical timbres be generalized to other types of timbre from other sound categories?

The American National Standards Institute [1] defined timbre as "that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different". However, another definition of timbre, which is "an invariant quality based on perceivable transformations across pitch and/or loudness that is assumed to underlie the ability to identify one instrument or voice" [7] has been used often in the psychoacoustic literature. We find this definition of timbre being more operational and useful for the present research, because each musical instrument can have its own timbre across different pitches and it extends the notion of timbre to other categories such as voice.

In order to find out if musicians have a superiority over non-musicians in processing timbre, Pitt [13] observed how the two groups perceived dimensions of timbre and pitch. He asked participants to tell if two consecutive musical tones were different in timbre and/or pitch. Participants had a four-choice categorization task: *no change, pitch change, instrument change,* and *both*

* Corresponding author. Tel.: +1 514 343 6111x1495; fax: +1 514 343 2285.
E-mail address: jean-pierre.chartrand@umontreal.ca (J.-P. Chartrand).
URL: http://vnl.psy.gla.ac.uk/index.html.

0304-3940/$ – see front matter
doi:10.1016/j.neulet.2006.06.053

changes. Results showed that when timbre changed, the musicians were far more accurate than the non-musicians. Because the participants had a four-choice task, they had to focus on both timbre and pitch variations on each trial. It has been shown that timbre variations can affect the judgement of pitch [9,15], especially in non-musicians [13]. It is possible that the non-musicians who participated in Pitt's study were more impaired than musicians when they had to judge timbre variations because they needed to concentrate on both dimensions simultaneously and thus were. That study only partially demonstrated that musicians do have a superiority over non-musicians in processing musical timbres, because it could have been a consequence of their superiority in pitch processing.

Münzer et al. [10] conducted three experiments to determine if musician's familiarity with tonal processing can be generalised to other kinds of superiority when processing other auditory features. In different recognition tasks, they found that musicians were better than non-musicians with musical instrument timbres, speech, and tones.

Speaker discrimination can be viewed as a particular case of timbre processing within a homogeneous category of sounds, and it is not known whether discriminating human voices and musical instruments involve different or similar processes. The present experiment asked the question of whether musical expertise with instrument timbres transfers to vocal timbres. In other words, will musicians be better than non-musicians at a vocal timbre discrimination task? We compared groups of musicians and non-musicians at two timbre discrimination tasks, one intrumental and one vocal. We predicted that musicians would perform better than non-musicians at both tasks, based on the results of Münzer et al. [10]. However, it is possible that musicians do not perform better than non-musicians since voice processing is a common and frequent task in both groups.

Thirty-six participants were recruited at the University of Montreal. Most of them were undergraduate students. The sample was composed of 17 musicians (9 women, 8 men) and 19 non-musicians (11 women, 8 men). A one-way ANOVA was performed to compare age differences between musicians (mean = 24.23, S.D. = 5.3) and non-musicians (mean = 23.94, S.D. = 5.93). The difference between groups was not significant ($F(1, 34) = 0.023$, $p > 0.05$). Musicians had at least 3 years of regular practice with an instrument or singing when included in this group (the years of training varied from 3 to 25 years). Out of the 17 musicians, there were 3 participants who were singers. When they were questioned about their musical training history, they reported that they also had courses with musical instruments, such as piano. Considering their specific training with voice production, they might be better than the other musicians at discriminating voices. However, the number of singers is so small that it would not make a large difference in the results. None of the musicians reported having absolute pitch. Both musicians and non-musicians reported having no auditory impairment. They participated on a voluntary basis and were not paid.

Two classes of stimuli were used: sounds of musical instruments and human voices. Thirty-two samples coming from four categories of musical instruments (piano, strings, guitar, brass) were created on a Roland JV-80 keyboard synthesizer. Each category included four instruments which played two sequences of notes: C-D-G and C-E-G. Those stimuli were recorded with Cooledit software (Syntrillium, 2000) in stereo and converted to mono with a sampling frequency of 22.05 kHz and a 16 bit resolution. The musical instrument samples had a mean length of 857 ms (S.D. = 27).

The human voices were taken from recordings of American vowels kindly provided by Hillenbrand [8]. The samples were also arranged in four categories: voices from women, men, boys, and girls. Each category contained four speakers pronouncing the syllables "had" and "heed". The samples were recorded in mono with a sampling frequency of 16 kHz and a 16 bit resolution. The mean length was 591 ms (S.D. = 87). Both classes of stimuli were normalized on mean energy (RMS) using Matlab (Mathworks).

Tasks were designed to be as similar as possible for the instrument and voice sounds. Sounds were presented in pairs with a 1 s inter-onset-interval. Half of the pairs came from the same source (same instrument, or same voice); but all the pairs differed on the spectro-temporal pattern, whether the source was the same or not: the two melodies of a pair were always different for the instruments task, and the two syllables of a pair were always different for the speakers task. The experiment was divided into 4 blocks of 96 pairs, 2 for the instruments and 2 for the voices. Response time was recorded from the onset of the sound. The next trial was initiated 2 ± 0.5 s after the response to the preceding pair. Accuracy was measured with the *d*-prime (d') measure of sensitivity [6].

Participants were set in a sound proof cabin in front of a computer keyboard. On each trial, a pair of stimuli was presented via Beyerdynamic DT 770 headphones. Instructions were exactly the same for the voice and instrument tasks: "Indicate with the keyboard if the two sounds presented are produced from the same sound source or not, while responding as fast as possible and maintaining the lowest error-rate possible"; i.e., same or different instrument, or same/different speaker.

The design of the experiment was 2×2 mixed factorial, with musical training (musician versus non-musician) as the between-participants factor and task (voice discrimination versus instrument discrimination) as the within-participants factor. Two ANOVAs were performed, in which the dependent variables were accuracy (d') and reaction time. Since there was no main effect of participant gender on reaction times ($F(1, 34) = 0.624$, $p > 0.05$) and discrimination performance ($F(1, 34) = 0.562$, $p > 0.05$), data were pooled across male and female participants. Mean response times of the participants were calculated for each task after removing responses that were two standard deviations above the mean for that participant were removed. Thus, corrected response time means were computed and used in the analyses. All the required ANOVA assumptions were checked during the analysis. Homogeneity of variance was equal across all categories (i.e., musicians versus non-musicians and voices versus musical instruments) for between-participant ANOVAs. In addition, Greenhouse–Geisser correction was used for repeated measures analysis to correct for potential violations of the sphericity assumption.

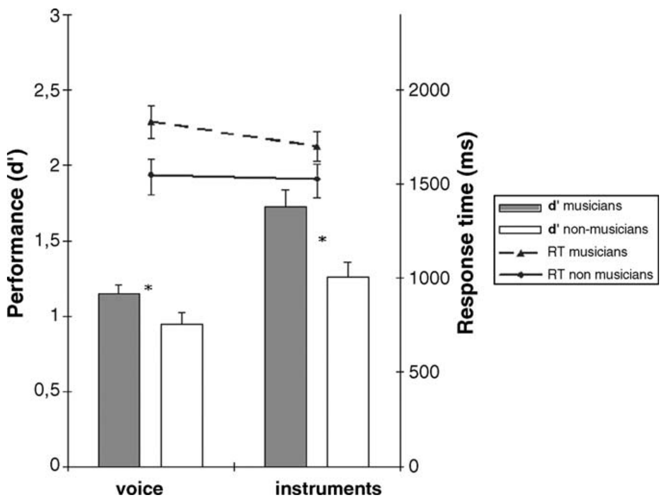

Fig. 1. Mean d' score (+S.E.) and response time for non-musicians and musicians in voice and instrument discrimination tasks.

There was a main effect of task on d' scores, (F(1, 34) = 31.482, $p < 0.05$). The voice discimination task was more difficult than the musical instruments discrimination task for all the participants. There was also a main effect of group: the musicians were found to perform significantly better than non-musicians at both the voice discrimination and the instrument discrimination tasks, (F(1, 34) = 10.834, $p < 0.05$). The interaction between musical training and task was almost significant, (F(1, 34) = 2.950, $p = 0.095$), suggesting that musicians had a greater advantage over the non-musicians in the instrument discrimination task. The results are shown in Fig. 1. A weak but significant correlation was found between voice and instrument discrimination performance ($r = 0.359$, $p < 0.05$). In order to eliminate outliers, individual ratios were computed and then transformed into z-scores. No participants had z-score ratios above or below 3.29 standard deviations from the mean ratio, so no participants were removed from the correlation. As shown in Fig. 2, the singing musicians had particularly good voice and instrument discrimination performance.

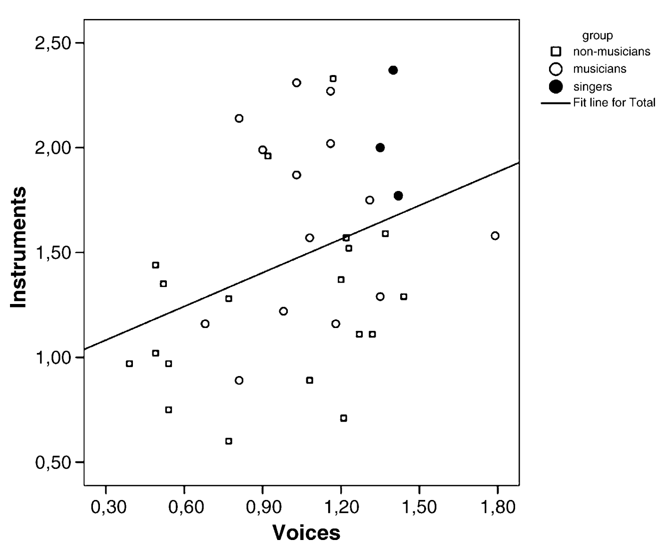

Fig. 2. Correlation between instruments and voices discrimination tasks performance (d') for both musicians and non-musicians.

Response times tended to be longer for musicians, (F(1, 34) = 3.497, $p = 0.07$). The musical training × task interaction was not significant, (F(1, 34) = 0.315, $p > 0.05$). The simple effect of musical training on the voice discrimination task was significant, (F(1, 34) = 4.16, $p < 0.05$), the musicians having significantly longer response times than the non-musicians (cf. Fig. 1).

As predicted, the musicians were found to perform better than the non-musicians at both voice and musical instrument discrimination tasks. The musicians took more time to respond in the voice discrimination task. Participants performed better at the instrument discrimination task than at the vocal discrimination task, but this.

One interpretation of the results is that expertise with musical instrument timbres generalizes to other timbre tasks, such as voice discrimination. Human voice is special in the sense that it is the carrier of speech and contains rich paralinguistic information about the speaker's identity, gender, emotional state [2]. But when it comes to auditory processing of acoustical features, it is unknown whether voice discrimination involves processes that are different from the ones implicated in instrument discrimination or not. Since there is a relation between the participants' musical education and timbre discrimination performance, we suppose that during their musical training, the musicians have learned to better discriminate timbres of various sound sources, but this remains to be further assessed. However, since the singing musicians seemed to be among the best performers in both discrimination tasks, experiments exploring the specific link between singing and timbre processing should be conducted.

Also, to accept this interpretation, we need to suppose that voice timbre and instrument timbre discrimination involve similar processes. If voice and instrument timbre processing were mediated by different processes, one would have expected the musicians to have only performed better at the instrument discrimination task. Musicians and non-musicians would have been expected to perform at the same level with the voice timbre discrimination task, since voice processing is a common domain of expertise across the whole population [2]. Not only did the musicians perform better than non-musicians at both tasks, but they showed a tendency to have a greater advantage in the instruments discrimination task. This result is not in direct contradiction with the proposal that the processes involved in the two tasks are similar. Considering the training that musicians have received with musical instruments, they were much more familiar than the non-musicians with the task involving instrument timbres. While the present study was limited to the question of timbre processing across different sound categories, it raises the question of functional modularity of timbre processing. Other experiments should be performed to examine whether the two types of timbre discrimination can interfere with each other (see Gauthier and Kurby [5], for an example of an interference task in visual perception).

An alternative explanation of the musicians' better performance is that they could have a greater non-verbal I.Q. This characteristic was not measured in our experiment. A study conducted by Brandler and Rammsayer [4] tested different aspects of mental abilities across musicians and non-musicians. It demonstrated that musicians were significantly superior only on auditory verbal memory and reasoning scales. It is true that a larger short-term auditory memory might explain why musicians outperformed non-musicians. Actually, auditory short-term memory was not measured in this study. However, if this explanation was true, then other studies would need to be conducted in order to assess whether this difference between musicians and non-musicians is caused by musical practice. Since little is known about musical instrument and voice timbre processing, for the moment it is difficult to attribute direct links between specific processes and good auditory discrimination performance. These results may partially explain our finding that the musicians had longer response times in both tasks.

The musicians' longer response times can be explained by arguments proposed by Münzer et al. [10]. In conclusion to their experiments, these authors said that musicians could process sound at a deeper level when they were given more time to encode the stimuli (168 and 200 ms versus 68 ms stimulus duration). The authors estimated that given more time for encoding, performance increases and an advantage for the musicians shows up. They proposed that musicians could have proceeded even further with the analysis of the auditory features. Although the participants were all instructed to respond as fast as possible while maintaining the lowest error-rate possible, the musicians took more time to respond overall and may have processed the sounds at a deeper level than the non-musicians, or used different cognitive strategies.

Finally, participants performed less well at the voice discrimination task than at the instrument discrimination task, but this piece of result is less relevant, since the objective of this study was to assess the link between musical training and timbre processing and thus to compare the two groups. The task demands were exactly identical in the two discrimination tasks, except for the timbre source. There were three differences though. While voices contained approximately one pitch during the whole syllable, musical instruments constituted of three different pitches. This may have lowered the difficulty of the instrument discrimination task. Also, some might argue that voice pairs differed in another aspect of timbre by having different vowels. This difference between musical instruments and voices could have easily led to a lower voice discrimination performance, since voice samples always differed in timbre, as opposed to musical instruments, which half of the time only differed in melodies. A difference in stimuli length can be noted too. The voice sounds had a mean total duration of 650 ms while the instrument sounds had a mean duration of 850 ms. The participants had 200 ms more of auditory information to process. Those two differences between voice and musical instrument sounds by themselves can explain why the instrument discrimination task led to better performance.

The main goal of the present study was to assess if musical timbre expertise can be generalized to timbres of other auditory domains, like voices. Our results suggest that it can, but this statement is not without limitations. First of all, it is not known yet if musicians are better than non-musicians at other voice processing tasks, like voice recognition for example. Second, it is also not known yet if musicians would perform better with other kinds of auditory stimuli, like environmental sounds. Future studies need to focus on multiple types of sound categories and other types of sound processing to further evaluate differences between muscians and non-musicians.

References

[1] ANSI (1973). Psychoacoustical terminology. S3.20. New York: American National Standards Institute.
[2] P. Belin, S. Fecteau, C. Bédard, Thinking the voice: neural correlates of voice perception, Trends Cogn. Sci. 8 (2004) 129–135.
[3] D.J. Bosnyak, R.A. Eaton, L.E. Roberts, Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones, Cerebr. Cortex 14 (2004) 1088–1099.
[4] S. Brandler, T.H. Rammsayer, Differences in mental abilities between musicians and non-musicians, Psychol. Music 31 (2003) 123–138.
[5] I. Gauthier, K.M. Curby, A perceptual traffic jam on highway N170, Curr. Dir. Psychol. Sci. 14 (2005) 30–33.
[6] D.M. Green, J.A. Swets, Signal Detection and Psychophysics, Wiley, New York, 1966.
[7] S. Handel, M.L. Erickson, A rule of thumb: the bandwith for timbre invariance is one octave, Music Percept. 19 (2001) 121–126.
[8] J. Hillenbrand, L.A. Getty, M.J. Clark, K. Wheeler, Acoustic characteristics of American English vowels, J. Acoust. Soc. Am. 97 (1995) 3099–3111.
[9] B.C.J. Moore, B.R. Glasberg, Frequency discrimination of complex tones with overlapping and non-overlapping harmonics, J. Acoust. Soc. Am. 87 (1990) 2163–2177.
[10] S. Münzer, S. Berti, T. Pechmann, Encoding of timbre, speech, and tones: musicians vs. non-musicians, Psychologische Beitrage 44 (2002) 187–202.
[11] C. Pantev, R. Oostenveld, A. Engelien, B. Ross, L.E. Roberts, M. Hoke, Increased auditory cortical representation in musicians, Nature 392 (1998) 811–814.
[12] C. Pantev, L.E. Roberts, M. Schulz, A. Engelien, B. Ross, Timbre-specific enhancement of auditory cortical representations in musicians, Neuroreport 12 (2001) 1–6.
[13] M.A. Pitt, Perception of pitch and timbre by musically trained and untrained listeners, J. Exp. Psychol. Hum. Percept. Perform. 20 (1994) 976–986.
[14] A. Shahin, D.J. Bosnyak, L.J. Trainor, L.E. Roberts, Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians, J. Neurosci. 23 (12) (2003) 5545–5552.
[15] P.G. Singh, I.J. Hirsh, Influence of spectral locus and F0 changes on the pitch and timbre of complex tones, J. Acoust. Soc. Am. 95 (1992) 2650–2661.
[16] L.J. Trainor, A. Shahin, L.E. Roberts, Effects of musical training on the auditory cortex in children, Ann. N. Y. Acad. Sci. 999 (2003) 506–513.