Grouping of Sequential Sounds—An Event-Related Potential Study Comparing Musicians and Nonmusicians

Titia L. van Zuijen[1], Elyse Sussman[2], István Winkler[3], Risto Näätänen[1], and Mari Tervaniemi[1]

Abstract

■ It is believed that auditory processes governing grouping and segmentation of sounds are automatic and represent universal aspects of music perception (e.g., they are independent of the listener's musical skill). The present study challenges this view by showing that musicians and nonmusicians differ in their ability to preattentively group consecutive sounds. We measured event-related potentials (ERPs) from professional musicians and nonmusicians who were presented with isochronous tone sequences that they ignored. Four consecutive tones in a sequence could be grouped according to either pitch similarity or good continuation of pitch. Occasionally, the tone-group length was violated by a deviant tone. The mismatch negativity (MMN) was elicited to the deviants in both subject groups when the sounds could be grouped based on pitch similarity. In contrast, MMN was only elicited in musicians when the sounds could be grouped according to good continuation of pitch. These results suggest that some forms of auditory grouping depend on musical skill and that not all aspects of auditory grouping are universal. ■

INTRODUCTION

The Gestalt theory originally formulated a description of how the visual system groups elements of the visual input into meaningful percepts (Koffka 1935). Gestalt theorists postulated that perceptual organization of the visual scene is guided by a set of principles. For example, the principle of similarity states that elements with similar features will be perceived as a unit separated from those that have less similar features. The principle of good continuation in turn states that visual elements following the same trajectory belong together. The Gestalt theory has more recently provided a framework for studying auditory perception (Deutsch, 1999; Bregman, 1990). In order to form meaningful auditory percepts it is especially important for the auditory system to structure ever-changing acoustic features and integrate sound over time. Gestalt principles can be used to describe temporal organization of sounds. According to the principle of similarity, consecutive sounds that contain similar features (e.g., pitch) would be grouped together and segmented from sounds with dissimilar features (e.g., another pitch). According to the principle of good continuation, series of sounds with smoothly changing features, such as notes following a melodic line, would be perceived as belonging together.

Originating from the Gestalt theory is the postulate that grouping processes of perceptual organization are automatic and universal. Automatic refers to the idea that attention is not required to link features of an object into a single coherent percept. Universality refers to the notion that organizational processes operate similarly regardless of the listener's age, culture, or musical skill (Drake & Bertrand, 2001, Imberty, 2000; Trehub, 2000). In line with these concepts of automaticity and universality, Jackendoff and Lerdahl (1983) proposed that perceptual groupings in music are intuitively formed and that the listener does not depend on musical knowledge in order to assign structure to a given passage of music (i.e., they are ''idiom independent''). Furthermore, the authors formulated grouping-preference rules for music perception derived from the Gestalt theory. Deliège (1987) and Peretz (1989) tested the empirical validity of the postulate that musical knowledge does not modify the usage of these grouping-preference rules. Musicians and nonmusicians were asked to divide musical excerpts taken from classical (Deliege, 1987) or folk music (Peretz, 1989) into segments. Their judgments were compared with the location of the grouping boundaries as predicted from the grouping-preference rules. The results of both studies were not completely in accordance with the hypothesis that musical expertise does not affect perceptual grouping, as it was found that musicians apply certain grouping rules more often than nonmusicians. Subjects were, however, attending to the sounds as they performed the task; therefore, task related top-down influences might have been at play, possibly affecting the basic perceptual organization of the sound material (Sussman, Ritter, &

[1]University of Helsinki, [2]Albert Einstein College of Medicine, [3]Hungarian Academy of Sciences

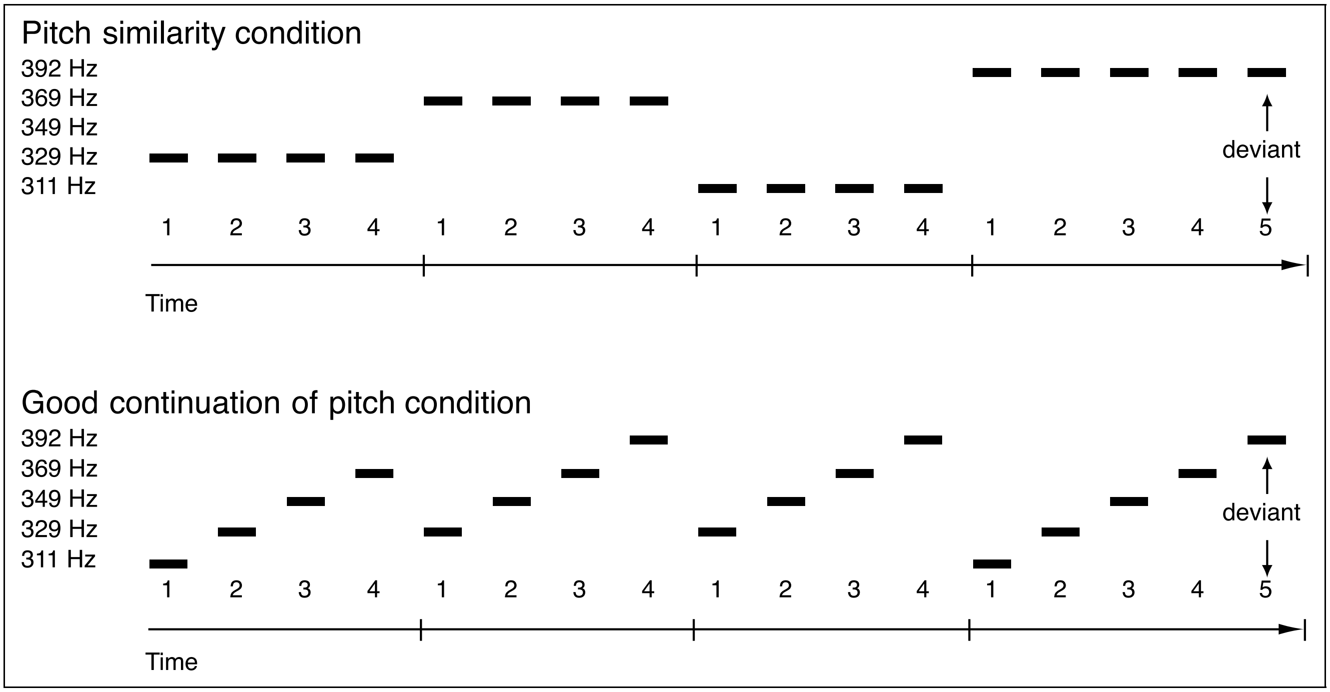

Figure 1. Schematic illustration of the paradigm. The upper panel displays the pitch similarity condition, in which four tones of one pitch are followed by four tones of another pitch (varying on five pitch levels). The deviant is a fifth tone violating the standard tone-group length by continuing in the same pitch as the preceding four tones. The lower panel displays the good continuation of pitch condition, in which four tones of a rising pitch follow each other (randomly starting on two pitch levels). The deviant is a fifth tone violating the standard tone-group length by rising one more pitch step.

Vaughan, 1998a; Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Peretz, 1989).

In the present study, we investigated whether auditory grouping processes differ between musicians and nonmusicians while eliminating the influence of task-related processes. For this purpose, we measured the mismatch negativity (MMN) of the auditory event-related potentials (ERPs) (Näätänen, Gaillard, & Mäntysalo, 1978; for reviews, see Picton, Alain, Otten, Ritter, & Achim, 2000; Schröger, 1997; Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995). The MMN reflects an auditory change detection process based on neural representations of acoustic repetitions or regularities. When occasionally a sound is encountered that does not match these memory representations because the repetition or regularity is broken an MMN is elicited. The MMN has been termed preattentive, as it can be elicited even when the subjects are not attending to the auditory stimuli; they can concurrently watch a movie, perform a visual, or even an auditory task with sounds delivered to the other ear (Alho, Woods, & Algazi, 1994; Alho & Sinervo, 1997). The memory representations underlying MMN elicitation can contain not only acoustic features such as pitch, duration, or location, but also more ecologically valid sounds such as vowels or relations between consecutive tones, like alternation, ascension, or a melody contour (Trainor, McDonald, & Alain, 2002; Tervaniemi, Rytkönen, Schröger, Ilmoniemi, & Näätänen, 2001; Paavilainen, Jaramillo, Näätänen, & Winkler, 1999; Nordby, Roth, & Pfefferbaum, 1988; for a review, see Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). In addition, MMN studies have shown that auditory grouping does not require attention, for instance, grouping by temporal proximity (Winkler, Schröger, & Cowan, 2001; Sussman, Ritter, & Vaughan, 1998b; Schröger, Tervaniemi, Wolff, & Näätänen, 1996) or by an interaction between temporal proximity and pitch similarity, as in the auditory streaming effect (Shinozaki et al., 2000; Sussman, Ritter, & Vaughan, 1999).

The MMN is generated irrespective of the attentional focus and is thought to reflect fairly automatic and basic auditory functions. MMN elicitation can nevertheless be affected by musical training. Koelsch, Schröger, and Tervaniemi (1999) found that an MMN was elicited in professional violinists by chords that deviated from repeatedly presented standards chords by only 0.75% in pitch, whereas musical novices showed an MMN only to much larger pitch deviations. Furthermore, Rüsseler, Altenmüller, Nager, Kohlmetz, and Münte (2001) found that compared with nonmusicians, musicians have a longer temporal span within which sound omissions elicit an MMN (Yabe, Tervaniemi, Sinkkonen, Huotilainen, Ilmoniemi, & Näätänen, 1998), suggesting that musicians can integrate sound over a longer period of time. Taken together, these studies provide evidence of enhanced preattentive auditory processing in musicians compared with nonmusicians, suggesting that fundamental auditory abilities, such as the preattentive encoding of spectral and temporal features can be enhanced in musical experts. Generalizing the finding that musicians have superior fundamental auditory processing abilities, we hypothesized that musicians may also have a more advanced ability to group sequential tones compared to nonmusicians. Finding that musicians demonstrate more advanced grouping abilities than nonmusicians would indicate that—despite its fundamental character—auditory grouping processes are not independent of musical skill.

We tested musicians and nonmusicians with two grouping rules, one based on pitch similarity and the other on good continuation of pitch. We expected that grouping based on pitch similarity would be easier than grouping based on extracting a consistent change of pitch and, thus, that both musicians and nonmusicians could group according to the first rule but only musicians according to the second more complex rule. Accordingly, it was expected that an MMN would be elicited by the deviants in musicians for both grouping rules but in the nonmusicians only for the pitch similarity rule.

In the ''pitch similarity'' condition, tone groups are defined by pitch repetition and the tone-group boundaries by a pitch change (see Figure 1, top). In the ''good continuation of pitch'' condition, tone groups are defined by pitch ascension and the tone-group boundaries by a descending pitch step (see Figure 1, bottom). Note that the tones are presented in an isochronous rhythm, so that there are no temporal grouping cues available. Four consecutive tones form the frequent standard-tone groups. The deviant is an occasional fifth tone that continues according to the grouping rule of that condition but violates the standard group length. The fifth (deviating) tone can only be detected and an MMN elicited if a neural template of the whole, standard tone group has been formed. MMN elicitation would therefore signify that without the subject's attention being required, the isochronal sequence is organized into tone groups. We recorded ERPs from professional classical musicians as well as from nonmusicians (see Methods). During the recording sessions subjects watched a silent subtitled video of their interest and were instructed to ignore the sound sequence.

RESULTS

Grouping Based on Pitch Similarity

In Figure 2, the grand-averaged ERPs recorded in the pitch similarity condition are shown for both musicians and nonmusicians. In the musicians, an MMN was elicited as indicated by the negative deflections in the waveforms for the deviant (fifth position tones) tones compared with the waveforms of the standard (fourth position tones): main effect of stimulus type, $F(1,10) = 66.3$, $p < .00001$, with significant effects at Fz, $F(1,10) = 36.8$, $p < .0001$; L1, $F(1,10) = 41.5$, $p < .0001$; R1, $F(1,10) = 37.9$, $p < .0001$; and pooled over the mastoid sites, $F(1,10) = 8.0$, $p < .02$. In the nonmusicians, an MMN was elicited as well: main effect of stimulus type, $F(1,11) = 21$, $p < .001$, with significant effects at Fz, $F(1,11) = 15.3$, $p < .003$); L1, $F(1,11) = 8.8$, $p < .01$); and R1, $F(1,11) = 15$, $p < .003$.

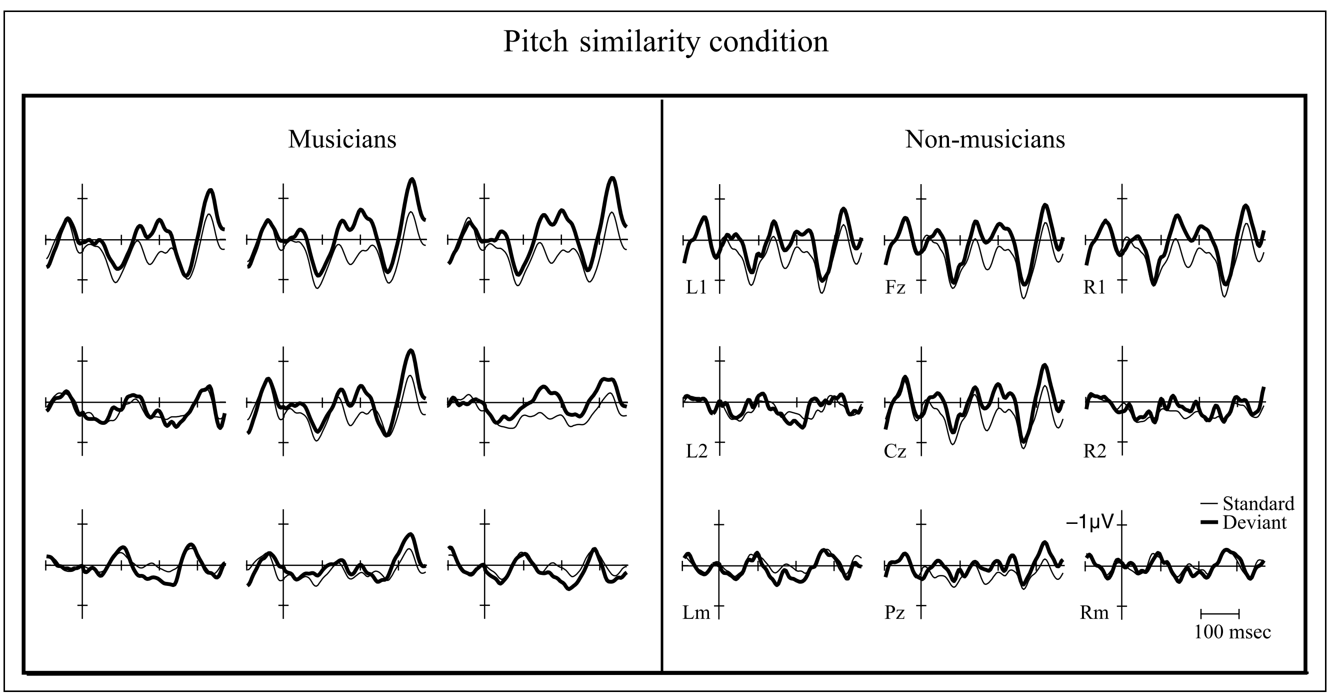

Figure 2. ERPs elicited in the pitch similarity condition. The grand-averaged ERPs are shown that were elicited in the pitch similarity condition by the standard (thin line) and deviant (thick line) tones at all channels for the musicians (left) and for the nonmusicians (right).

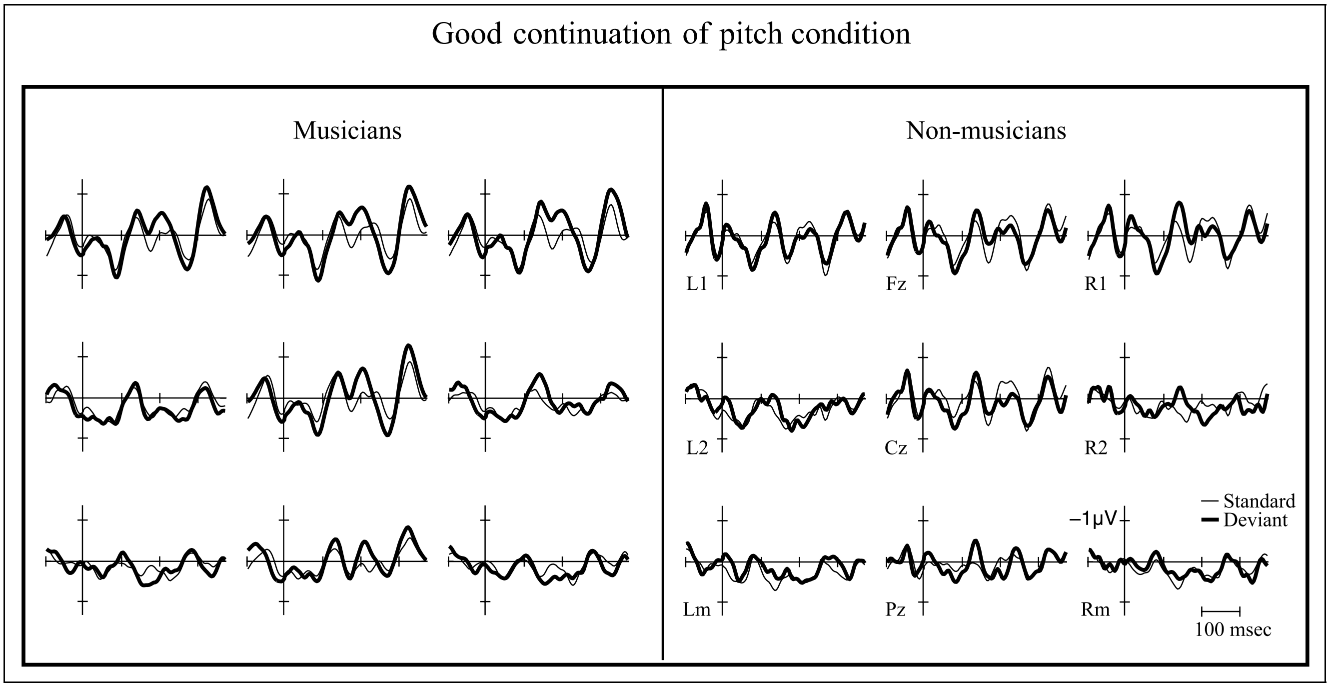

Figure 3. ERPs in the good continuation of pitch condition. The grand-averaged ERPs are shown that were elicited in the good continuation of pitch condition by the standard (thin line) and deviant (thick line) tones at all channels for the musicians (left) and for the nonmusicians (right).

Grouping Based on Good Continuation of Pitch

Figure 3 shows the grand-averaged ERPs recorded in the good continuation of pitch condition. An MMN was elicited in the musicians, as indicated by the negative difference between the ERPs elicited by deviant and standard tones: main effect of stimulus type, $F(1,10) = 29.0, p < .0003$, with significant effects at Fz, $F(1,10) = 9.8$, $p < .01$; L1, $F(1,10) = 6.4, p < .03$; and R1, $F(1,10) = 10.3$, $p < .009$.

Differences between Levels of Expertise and between Grouping Rules

Figure 4 gives an overview of the ERPs elicited in the two subject groups and the two conditions. The ANOVA on the MMN amplitude (the deviant minus standard difference) for level of expertise and conditions yielded a main effect of expertise, $F(1,21) = 5.1, p < .04$, indicating that the MMN of the musicians was overall larger than the MMN of the nonmusicians. The subsequent post hoc comparisons showed a trend towards the MMN being larger in musicians than in nonmusicians in both the pitch similarity, $F(1,21) = 3.2, p < .09$, and the good continuation of pitch condition, $F(1,21) = 3.8, p < .06$. Furthermore, a main effect of condition was obtained, $F(1,21) = 6.9$, $p < .02$, as the MMN amplitude was overall larger in the pitch similarity condition compared to the good continuation of pitch condition. The post hoc comparisons showed that this was mainly caused by the nonmusicians whose MMN in the pitch similarity condition was larger than the deviant minus standard difference in the good continuation of pitch condition, $F(1,21) = 5.2$, $p < .03$, whereas in musicians the MMN amplitudes did not significantly differ across conditions.

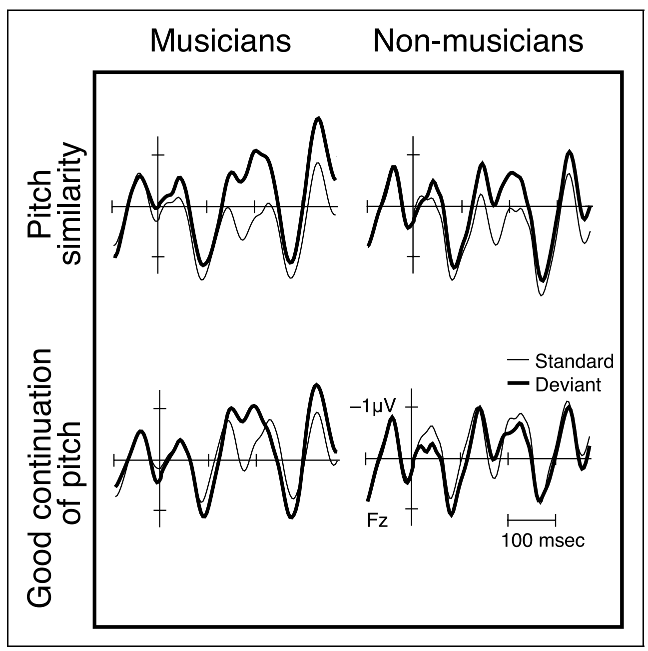

Figure 4. Overview of ERPs. The waveforms elicited at Fz re-referenced to the average of the two mastoids are presented for the two subject groups and the two conditions. The upper panel shows the waveforms that were elicited in the pitch similarity condition; the lower panel shows the waveforms that were elicited in the good continuation of pitch condition.

The interaction between expertise and condition was, however, not significant.

DISCUSSION

The aim of this study was to determine whether fundamental auditory grouping abilities depend on musical expertise, which could indicate that auditory grouping is not solely universal. To this end, we tested whether musicians and nonmusicians could group an isochronous tone sequence into segments of four consecutive tones by using two different grouping rules: one based on the Gestalt principle of similarity, the other based on the principle of good continuation. Grouping was indexed by the elicitation of the MMN to occasional deviant tones that violated the possibly emerging tone groups.

Our hypotheses predicted that musicians would show an MMN in both conditions, but the nonmusicians in the pitch similarity condition only. Accordingly, we found an MMN for the musicians and for the nonmusicians in the pitch similarity condition, indicating that grouping based on pitch similarity took place irrespective of level of expertise. This suggests that certain aspects of sequential grouping operate irrespective of musical expertise. This is in concurrence with previous findings showing auditory grouping for task-irrelevant sounds (Winkler et al., 2001; Sussman et al., 1999; Schröger et al., 1996). We also found an MMN, for the musicians, in the good continuation of pitch condition indicating that musicians could organize the tones according to the changing pitch relations. In contrast, there was no evidence of an MMN in the nonmusicians; taken together with the finding that nonmusicians could group by pitch similarity this suggests that the additional complexity of the good-continuation principle hindered them to organize the tone sequence into four-tone groups.

The MMN of the two subject groups did not differ depending on condition. Finding such an interaction would have indicated that musicians are specifically superior in grouping based on good continuation of pitch compared to nonmusicians, whereas the results showed a larger deviant minus standard difference for the musicians in both conditions.

Musicians and nonmusicians may differ in auditory grouping in two ways. On one hand, they might differ in a quantitative way if grouping processes operate generally similarly in musicians and in nonmusicians with the only difference that musicians are able to form stronger group associations. When a grouping rule is violated by a deviant this would be more salient for the musicians than for the nonmusicians. In this case, an MMN would be expected to be elicited in both subject groups and in both conditions, but it would, however, trigger a larger response in the musicians than in the nonmusicians. Our results are compatible with this interpretation if we assume that we did not find evidence for an MMN in the nonmusicians in the good continuation of pitch condition because it was too small to be detected. Given that assumption the MMN of the musicians was overall larger than the MMN of the nonmusicians. On the other hand, grouping processes may operate qualitatively differently in musicians than in nonmusicians; in which case musicians can form group associations that nonmusicians cannot form. Our data are compatible with this interpretation if we assume that there was no evidence for an MMN in the good continuation of pitch condition for the nonmusicians because they did not elicit any MMN. This interpretation entails that musicians are able to apply a larger variety of rules for structuring sound sequences than individuals who did not reach this level of expertise in music. To conclude, the differences that we have found between musicians and nonmusicians in auditory grouping indicate that auditory grouping is not independent of musical skill. Furthermore, if one adheres to the latter interpretation that musicians and nonmusicians differ in a qualitative way this would then suggest that auditory grouping is not universal.

The observed difference between musicians and nonmusicians in preattentive sequential auditory grouping may be a result of the functional importance for musicians to retain and recognize melodic patterns. It is possible that musicians, compared to nonmusicians, have more accurate neural representations of intervallic relationships, resulting in more advanced neural templates of multiple-tone segments, such as prototypical melodic patterns, stored in their long-term auditory memory (Oura, 1991). This is compatible with previous findings of interactions between long-term memory for speech sounds and the transient memory involved in the MMN generation (Schröger, 2000; Winkler, Cowan, Csépe, Czigler, & Näätänen, 1996). For example, a vowel belonging to a native vowel category, presented as a deviant among other native vowels, elicited a larger amplitude MMN compared to a deviant vowel that fell outside the native vowel categories of the subject (Winkler et al., 1999; Näätänen et al., 1997; for a review, see Näätänen, 2001). In a similar fashion, more accurate templates of tonal patterns stored in the long-term memory might interact with the ability to preattentively group/structure consecutive sounds.

More advanced preattentive grouping in musicians may be a result of a shift from controlled to automatic processing due to training (Jansma, Ramsey, Slagter, & Kahn, 2001; Shiffrin & Schneider, 1977).[1] That musical training can lead to neural plasticity in general is indicated by a positive correlation between the starting age of musical training and an enhanced response to musical sounds as compared to pure tones (Pantev et al., 1998), as well as by a response that is specifically enhanced to tones played by the instrument of training of professional musicians (Pantev, Roberts, Schulz, Engelien, & Ross, 2001). That training can specifically modify preattentive processing has been shown in both nonmusical (Atienza,

& Cantero, 2001; Menning, Roberts, & Pantev, 2000; Kraus, McGee, Carrell, King, & Tremblay, 1995; Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993) and musical subjects (Tervaniemi et al., 2001). In these studies, deviants are used that are difficult to detect and to which the MMN is initially absent. In a recording session following a short training, the MMN appears in those subjects that successfully learned to detect the deviants. These studies show that attention is needed for learning, but that once a new skill has been acquired it can be available without the requirement of attention (Näätänen et al., 1993). Extensive and finely tuned automatic processes in professional musicians could be beneficial for improving overall processing efficiency in their performance as this would leave more of the limited attentional resources available for higher level processes needed to perform music at a professional level.

METHODS

Subjects

Eleven musicians (age 22 to 28 years, mean 24, 4 men) and 12 nonmusicians (age 19 to 26 years, mean 22.4, 4 men) participated in the experiment. Musicians were defined as having reached, as a minimum, the level of acceptance into a music academy where they were to become classical performing artists. The musicians had started playing an instrument between 4 and 8 years of age (mean starting age was 5.6 years) and the average number of years of playing at a professional level was 4, with a daily study of, on average, 4.2 hr. Nonmusicians were defined as never having studied any form of music at a formal professional level. All but two of the nonmusicians played an instrument as an amateur. None of the nonmusicians played an instrument on a daily basis at the time of the experiment. All subjects were naive with respect to the paradigm and were paid for their participation.

Stimuli and Procedure

Pitch Similarity Condition

One-hundred-millisecond sine wave tones (50 dB above hearing threshold; 10-msec rise and 10-msec fall times) were presented with a constant intertone interval of 87.5 msec. Stimulus sequences consisted of monotonous (identical pitch) four-tone segments varying on five frequency levels, ranging from 311.1 to 392 Hz in semitone steps on the musical scale. Four-tone segments of 750 msec formed the standard tone groups. The tone-group boundaries were indicated by a change in pitch. Ten percent of these four-tone groups were prolonged by a fifth tone of the same pitch violating the standard group length (see Figure 1, upper panel).

The standard ERP was averaged from the responses evoked by the tones in the fourth positions. The deviant ERP was averaged form the responses evoked by the tones in the fifth (deviating) position.

Good Continuation of Pitch Condition

This condition consisted of the same five sine wave tones and the same intertone interval as the pitch similarity condition (see above). Groups of four tones were formed by ascending pitch steps, after which the onset of the next tone group was indicated by a falling pitch (Figure 1, lower panel). An ascending four-tone segment could start on two pitch levels (either starting at 311.1 Hz and ending at 369.9 Hz or starting one semitone step higher and ending at 392 Hz). Ten percent of the tone segments contained an additional ascending step. This was always a continuation of the tone group starting at the lowest pitch level in order to avoid the introduction of a new (and therefore deviating) frequency in the stimulus sequence. Furthermore, the deviant tone segment was always preceded by at least one standard ending at the highest (392 Hz) frequency (as in the example in Figure 1) to avoid an MMN being elicited by the possible rare local occurrence of the highest frequency.

The standard waveform was averaged from the responses to the highest frequency tones when they occurred at the end of a standard group. Only those standard groups were included that were preceded by at least one standard of the same kind because the deviants were also always preceded by such a standard. Responses elicited by the fifth-position tones were averaged together and are denoted as the deviant.

Procedure

The subjects were seated in a comfortable chair. They were instructed to watch a subtitled movie without sound and to disregard the auditory stimuli. The tone sequences were presented binaurally through headphones. Two blocks of stimuli of both conditions were presented in counterbalanced order. Each block lasted for approximately 10 min and contained 80 deviants. For each condition, 160 deviants were presented in total.

After the recording sessions, subjects were asked whether they had had any difficulties ignoring the stimuli. None of the subjects reported such difficulty, musicians and nonmusicians alike.

ERP Recording

Electroencephalogram (EEG) was recorded with Ag/AgC1 electrodes placed at three midline positions, Fz, Cz, and Pz, at the left and right mastoids (Lm and Rm, respectively) and at sites along the coronal chain at one third (L1 and R1) and two thirds (L2 and R2) between Fz and the mastoids, on both sides of the head. The reference electrode was placed at the tip of the nose.

The horizontal electrooculogram (EOG) was monitored using a bipolar configuration with electrodes placed lateral to the outer canthi of the two eyes. Vertical EOG was recorded with electrodes placed above and below the left eye. The signal was amplified between 0.5 and 50 Hz and digitized (Synamps amplifiers) at 250 Hz.

Data Analysis

The EEG was digitally filtered off-line between 1.5 and 35 Hz. Epochs starting 100 msec before and ending 375 msec after the onset of the tones were taken from the continuously recorded EEG. Baseline correction was applied on single trials after which trials that contained electrical activity exceeding ±75 μV at any electrode were rejected. On average, 25% of the deviant trials were rejected (out of the 160 presented). ERPs were averaged separately for each stimulus type (standard and deviant, see above) and condition.

A typical nose-referenced MMN emerges as an enhanced negativity of the deviant with respect to the standard waveform at frontal electrode sites (e.g., Fz, L1, and R1) and, usually, but not necessarily, a polarity reversal at the mastoid sites. To assess MMN elicitation, a two-way ANOVA was performed separately for each condition with repeated measures factors Electrode [Fz, L1, R1, Lm (inverted), Rm (inverted)] and Stimulus Type (standard, deviant). The measures obtained at the mastoids were inverted in polarity to avoid the cancellation of effects due to the mastoid polarity reversal. Planned post hoc comparisons were used to determine significant differences on individual electrodes. Amplitude measures were calculated by taking the mean amplitude separately at each channel in a 40-msec window centered on the peak latency of the grand-averaged deviant minus standard difference waveforms at Fz.

To assess differences between subject groups and conditions, an additional two-way ANOVA of the deviant minus standard difference amplitude was performed with the factor Expertise (musicians, nonmusicians) and the repeated-measure condition (pitch-similarity, good continuation of pitch). Planned post hoc comparisons were used to determine significant differences between conditions for each level of expertise and between levels of expertise for each condition. A measure of the MMN amplitude was calculated as the mean amplitude in a 40-msec window centered on the MMN peak latency from the difference waveform (deviant minus standard) at Fz, which was re-referenced to the average of the mastoids to include the polarity reversal at the mastoid leads.

In all statistical tests the alpha level was set at $p < .05$.

Acknowledgments

This research was supported by the Academy of Finland, the Finnish Graduate school in Psychology, the National Institutes of Health (Grant R01 DC04263), and the Hungarian National Research Fund (OTKA T034112). We thank Klaus Linkenkaer-Hansen for his helpful comments on the manuscript.

Reprint requests should be sent to Titia van Zuijen, Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, P.O. Box 13, FIN-00014, Finland, or via e-mail: titia@cbru.helsinki.fi.

Note

1. Talent or an interaction between talent and training may as well play a crucial role, but this discussion reaches beyond the scope of this article.

REFERENCES

Alho, K., & Sinervo, N. (1997). Preattentive processing of complex sounds in the human brain. *Neuroscience Letters, 233,* 33–36.

Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology, 31,* 469–479.

Atienza, M., & Cantero, J. L. (2001). Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Research, 901,* 151–160.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge: MIT Press.

Deliège, I. (1987). Grouping conditions in listening to music. An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception, 4,* 325–360.

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music.* (2nd ed., pp. 299–348) New York: Academic Press.

Drake, C., & Bertrand, D. (2001). The quest for universals in temporal processing in music. *Annals of the New York Academy of Sciences, 930,* 17–27.

Imberty, M. (2000). The question of innate competencies in musical communication. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music.* (pp. 449–462). Cambridge: MIT Press.

Jackendoff, F., & Lerdahl, R. (1983). *A generative theory of tonal music.* Cambridge: MIT Press.

Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *Journal of Cognitive Neuroscience, 13,* 730–743.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior attentive and pre-attentive auditory processing in musicians. *NeuroReport, 10,* 1309–1313.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Hartcourt Brace.

Kraus, N., McGee, T., Carrell, T., King, C., & Tremblay, K. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neurosciences, 7,* 27–34.

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport, 4,* 817–822.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MNN) and its magnetic equivalent (MMNm). *Psychophysiology, 38,* 1–21.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42,* 313–329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport, 4,* 503–506.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, R., & Winkler, I. (2001). ''Primitive intelligence'' in the auditory cortex. *Trends in Neurosciences, 24,* 283–288.

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology, 25,* 262–268.

Oura, Y. (1991). Constructing a representation of a melody: Transforming melodic segments into reduced pitch patterns operated on by modifiers. *Music Perception, 9,* 251–266.

Paavilainen, P., Jaramillo, M., Näätänen, R., & Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters, 265,* 179–182.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392,* 811–814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12,* 169–174.

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychophysiology, 24,* 157–178.

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, *5,* 111–139.

Ritter, W., Deacon, D., Gomes, H., Javitt, D. C., & Vaughan, H. D., Jr. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: A review. *Ear & Hearing*, *16,* 52–67.

Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., & Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neuroscience Letters, 308,* 33–36.

Schröger, E. (1997). On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology, 34,* 245–257.

Schröger, E. (2000). Top-down effects on auditory sensory memory processing. In A. Schick, M. Meis, & C. Reckhardt (Eds.), *Contributions to psychological acoustics, results of the eight Oldenburg symposium on psychological acoustics* (pp. 337–353). Oldenburg: Bibliotheks- und informationssystem der universitäts Oldenburg.

Schröger, E., Tervaniemi, M., Wolff, Ch., & Näätänen, R. N. (1996). Preattentive periodicity detection in auditory patterns as governed by time and intensity information. *Cognitive Brain Research, 4,* 145–148.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and general theory. *Psychological Review, 84,* 127–190.

Shinozaki, N., Yabe, H., Sato, Y., Sutoh., T., Hiruma, T., Nashida, T., & Kaneko, S. (2000). Mismatch negativity (MMN) reveals sound grouping in the human brain. *NeuroReport, 11,* 1597–1601.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998a). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research, 789,* 130–138.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998b). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport, 9,* 4167–4170.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology, 36,* 22–34.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research, 13,* 393–405.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J, & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory, 8,* 295–300.

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14,* 430–442.

Trehub, S. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 427–448). Cambridge: MIT Press.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, A. H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive change detection utilizes both phonetic and auditory memory representations in across- and within-category vowel contrasts. *Cognitive Brain Research, 7,* 357–369.

Winkler, I., Schröger, E., & Cowan, N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience, 13,* 59–71.

Winkler, I., Cowan, N., Csépe, V., Czigler, I., & Näätänen, R. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. *Journal of Cognitive Neuroscience, 8,* 403–415

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). The temporal window of integration of auditory information in the human brain. *Psychophysiology, 35,* 615–619.