

Evidence for Linear but Not Helical Automatic Representation of Pitch in the Human Auditory System

Tamar I. Regev, Israel Nelken, and Leon Y. Deouell

Abstract

■ The perceptual organization of pitch is frequently described as helical, with a monotonic dimension of pitch height and a circular dimension of pitch chroma, accounting for the repeating structure of the octave. Although the neural representation of pitch height is widely studied, the way in which pitch chroma representation is manifested in neural activity is currently debated. We tested the automaticity of pitch chroma processing using the MMN—an ERP component indexing automatic detection of deviations from auditory regularity. Musicians trained to classify pure or complex tones across four octaves, based on chroma—C versus G (21 participants, Experiment 1) or C versus F# (27, Experiment 2). Next, they were passively exposed to MMN protocols designed to test automatic detection of

height and chroma deviations. Finally, in an “attend chroma” block, participants had to detect the chroma deviants in a sequence similar to the passive MMN sequence. The chroma deviant tones were accurately detected in the training and the attend chroma parts both for pure and complex tones, with a slightly better performance for complex tones. However, in the passive blocks, a significant MMN was found only to height deviations and complex tone chroma deviations, but not to pure tone chroma deviations, even for perfect performers in the active tasks. These results indicate that, although height is represented preattentively, chroma is not. Processing the musical dimension of chroma may require higher cognitive processes, such as attention and working memory. ■

INTRODUCTION

Auditory pitch is a perceptual property of many sounds. The physical property most strongly associated with pitch perception is temporal periodicity of sound waves. Although the official ANSI definition of pitch is “that auditory attribute of sound, according to which sounds can be ordered on a scale from low to high” (ANSI, 1994), researchers frequently describe the perceptual organization of pitch as a two-dimensional helix (e.g., Moerel, De Martino, Santoro, Yacoub, & Formisano, 2015; Briley, Breakey, & Krumbholz, 2013; Warren, Uppenkamp, Patterson, & Griffiths, 2003; Wright, Rivera, Hulse, Shyan, & Neiwirth, 2000; Shepard, 1982). One dimension of the helix is pitch height (termed here simply as “height”)—a monotonic dimension constantly increasing when we move, for example, from left to right on the piano keyboard, as the period of the sound decreases. The second dimension is pitch chroma (chroma)—a circular dimension reflecting the repeating structure of the octave. If the periods of two sounds have a ratio of 2^n , where n is an integer, these sounds belong to the same chroma, and they are spaced exactly n octaves apart. In western music notation, the octave is divided to 12 pitch

classes, and all pitches within the same class (separated by an integer number of octaves) have the same chroma. For example, all tones belonging to the pitch class C have the same chroma, which we simply call C.

The reason for suggesting the helical model was behavioral evidence that tones with the same chroma are rated as perceptually similar, a phenomenon also known as the “octave equivalence” property (Hoeschele, Weisman, & Sturdy, 2012). When a group of people sings together, males usually sing an octave lower than females, but the result sounds in tune, because everyone sings in the same chroma. Octave generalization effects were shown behaviorally in nonmusicians using pure tones (Hoeschele et al., 2012). Even infants judge pure tone melodies spaced an octave apart as more similar than other interval transpositions (Demany & Armand, 1984) and Wright et al. (2000) provide evidence for octave generalization in Rhesus monkeys. However, although much is known about the manifestation of height in neural activity, the level at which chroma is manifested in the brain is debated. Several recent studies postulated that the neural organization of pitch is consistent with the helical model, providing evidence for chroma processing in the human cortex (e.g., Moerel et al., 2015; Briley et al., 2013; Warren et al., 2003). Although, as mentioned, there is some evidence that octave equivalence

can be detected in infants and rhesus monkeys, it is still not clear whether chroma is automatically processed or whether chroma processing requires higher order cognitive processes, such as attention.

The aim of this study was to test the automaticity of chroma processing in the human brain. We asked whether the processing of chroma is automatic and preattentive, as is well known for height. We used the MMN ERP to operationalize the notion of automatic processing, because the MMN is believed to index automatic detection of deviations from auditory regularity even in an ignored sound stream. We therefore checked whether deviations from chroma regularity elicit the MMN. In two EEG experiments, including 48 musicians (21 in Experiment 1 and 27 in Experiment 2), we hypothesized that if chroma is processed automatically, violations of chroma regularity will evoke an MMN.

EXPERIMENT 1

In the first experiment, we concentrated on the ability to discriminate between the pitch classes C and G. This pair of notes forms the “perfect fifth” interval that is usually the first one learned in “ear training” programs.

Methods

Participants

Twenty-five healthy musicians participated in the experiment and were paid 40 shekels (~US\$12) per hour. Participants were recruited either from the Jerusalem Academy of Music and Dance or from the Hebrew University. The criterion for inclusion in the experiment was at least 5 years of formal music training and active involvement with music today, either as students at the academy or professionally. The data of four participants were excluded because of a technical problem in the recording. The analysis included the data of 21 participants (seven women, mean age = 29 years, $SD = 8.9$ years). All participants self-reported normal hearing and no history of neurological disorders. Five participants reported having absolute pitch (AP). The experiment was approved by the ethical committee of the faculty of social science at the Hebrew University of Jerusalem, and informed consent was obtained after the experimental procedures were explained.

Stimuli and Apparatus

Participants were seated in a dimly lit, sound-attenuated, and echo-reduced chamber (C-26, Eckel) in front of a 17-in. CRT monitor (100-Hz refresh rate), at a viewing distance of about 90 cm. The screen was concealed by a black cover, with a rectangular window in the middle (14 × 8.5 cm), through which they viewed the visual display. Auditory stimuli were presented through earphones

(Sennheiser HD25, having a relatively flat frequency response function in the range of frequencies used in the experiment). The experiment was run using the Psychophysics toolbox (Brainard, 1997) for MATLAB (Version 2013b, MathWorks) running on a 32-bit system (Windows XP). Auditory stimuli were synthesized using MATLAB software. Experiment 1 included only pure tones, each of 100 msec duration with 30-msec-long linear rise and fall gates. Stimuli were presented at a sound pressure level that was comfortable for the participants. In the beginning of the experiment, the relative amplitude of each tone was adjusted such that participants reported equal loudness subjectively.

Experiment Design

Ear training. Participants classified tones according to their pitch chroma. The task was inspired by an ear training method developed by Dr. Bat-Sheva Rubinstein (Buchmann-Mehta School of Music, Tel Aviv University). Eight tones spanning four octaves were presented: four Cs (261.6, 523.2, 1046.5, and 2093 Hz, corresponding to C₄, C₅, C₆, and C₇, respectively) and four Gs (392, 784, 1568, and 3136 Hz, corresponding to G₄, G₅, G₆, and G₇, respectively). Each time that a tone was played, participants had to classify it according to chroma, assisted by singing either C or G in their preferred octave. The participants were instructed to sing the correct pitch and, in addition, to say the name of the note. Singing in the correct pitch was suggested to the participants as helpful for performance. However, when analyzing the results, only the name of the note was taken as the indication for a correct or incorrect answer. This was explained to participants. A short (unanalyzed) practice block was presented at the beginning of the session to let participants get acquainted with the task and stimuli. Responses were manually marked by the experimenter and recorded for later verification. During tone presentation, participants were asked to direct their eyes toward a small black fixation cross appearing on a gray background in the center of the screen. After each response, the correct answer replaced the fixation cross, as a black letter—C or G—according to the correct chroma of the notes. The notes appeared randomly, with 60% Cs and 40% Gs. There were four blocks, consisting of 50 tones each. Participants were informed that every block began with three Cs from different octaves. Beginning the sequence with three Cs, as well as assigning a larger probability to Cs over Gs, was used to base the tonic (Aldwell & Cadwallader, 2018) as C, that is, to keep a stable context of the C major key and avoid shifts to a G major context, which could confuse participants and affect the EEG (Poulin-Charronnat, Bigand, & Koelsch, 2006). The first three blocks were slower, with an SOA (i.e., the time passing between the onsets of two consecutive stimuli) of 3 sec in total, leaving 1.7 sec to respond and 1.3 sec with the correct answer displayed. The last block was

faster, with a total SOA of 1.5 sec, leaving 1 sec to respond and 0.5 sec for display of the correct answer.

Passive MMN. Participants viewed a silent movie. They were instructed to ignore stimuli presented through earphones and concentrate on the film. They could choose one of two movies—“The Kid” (Charlie Chaplin, 1921) or “Spirit: Stallion of the Cimarron” (Dreamworks LLC, 2002). There were four MMN block types (see Figure 1 for a schematic summary): height deviation, height control, chroma deviation, and chroma control. The height deviation block contained 80% standard tone D5 (587.3 Hz) and 20% deviant tone A5 (880 Hz). The deviant appeared in a pseudorandom order such that there were a minimum of three and a maximum of eight repeating standards between each two deviants. The height control block included five tones; Eb4, Db5, A5, F6, and B6 (311.1, 554.4, 880, 1397, and 1975.5 Hz, respectively) each presented 20% of the times in a pseudorandom order: Random permutations of the five tones were concatenated while ensuring that there were no repetitions and at least three other tones between each A5 presentation. The A5 tone in this sequence served as the

control tone to be contrasted with the deviant A5 in the height deviation block, because it was the same tone and appeared with the same probability in the control sequence, but among tones that did not form any regular pattern (as in Jacobsen & Schröger, 2001). The chroma deviation block included five tones—four standard tones having chroma C, from four octaves, C4, C5, C6, and C7 (261.6, 523.2, 1046.5, and 2093 Hz, respectively; the same tones that appeared in the ear training part), and the deviant tone was G5 (784 Hz). The five tones appeared each 20% of the times, such that none of them was a deviant by pitch height on its own. We reasoned that if chroma C is represented automatically in the brain, then the Cs would be grouped to form an 80% standard group and the G a 20% chroma deviant. The five tones were presented pseudorandomly in the same way as for the height control block. Finally, the chroma control block included the same G5 (784 Hz) tone included as deviant in the chroma deviation block, but the other four tones in this block were of four different chromatic values, namely Db4, B4, Eb6, and A6 (277.2, 493.9, 1244.5, and 1760 Hz, respectively). As a result, the middle tone G5 served as a comparable

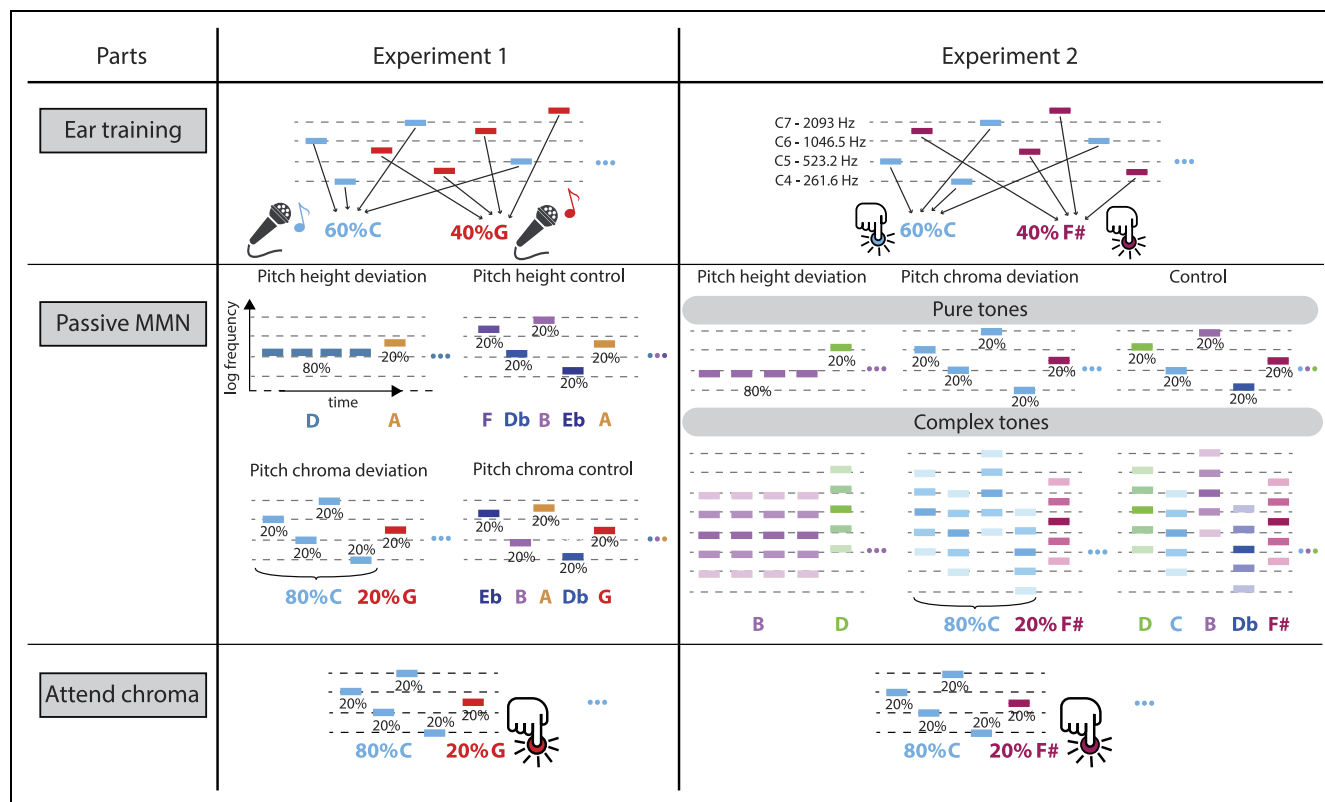


Figure 1. Summary of stimuli and design of Experiments 1 and 2. The three experimental parts are presented: ear training, passive MMN, and attend chroma. The dashed lines represent the auditory frequencies of C4–C7 (frequencies are specified in the ear training part of Experiment 2) on a logarithmic scale. Each colored line represents a pure tone having a specific frequency. The color represents the chroma of the tone. The pitch class in western music notation is specified under tones with the same color (e.g., C for light blue, G for red, etc.) Lighter colors in the complex tones part of Experiment 2 represent tones with lower intensity. Each sequence of tones is represented schematically by five representative tones, and the probability of presentation of each tone along the sequence is denoted as percentage. The button and microphone icons represent the task (detection by pressing a button or by singing, respectively). See Methods (Stimuli and Apparatus and Experiment Design sections) for further detail.

control for the G5 deviant tone in the chroma deviation block (see Figure 1 for an illustration). Twelve MMN blocks in total were presented, three from each type. Each block included 500 trials, 100 of each specific tone (for the pitch height deviation block, there were 400 standards and 100 deviants). This resulted in 300 trials for each deviant and for its comparable control. The tones were presented with an SOA of either 450 or 550 msec, randomly (average SOA was 500 msec). As a result, each block took 250 sec, and there was a 30-sec break between the blocks, unless the participant asked for more time. In total, this part required about an hour of EEG recording. Because it was hard for the participants to stay alert for such a long time, we recorded four blocks before the ear training task, and the other after. For the early recordings, we selected the blocks that did not interact with this task—the height deviation and height control. As a result, the order of the blocks was counterbalanced between the participants in the following way. Denoting the height deviation block, the height control block, the chroma deviation block, and the chroma control blocks by A, B, C, and D, respectively, half of the participants were tested in the order ABAB before the ear training part and CDCDBADC after, and the other half were tested in the order BABA before the ear training part and DCDCABCD after.

Attend chroma. The same stimuli as in the chroma deviation MMN block were presented (C4, C5, C6, C7, and G5) with the same probabilities and randomization procedure, and participants had to press a button each time that the target G appeared. This was done to verify that participants could detect the deviant tone and to assess their performance in a similar setting to the passive MMN block. No feedback was given on button presses. Because an SOA of 500 msec was too short for pressing a button before the next tone, the SOA here was longer, either 1300 or 1700, randomly. Two blocks of 150 trials each were presented with a short rest between them. This resulted in a total of 60 target presentations. To avoid interfering with the main task, the monitor displayed a frozen frame from the movie.

To summarize, participants started with four blocks of the passive height and control (~20 min), then performed the ear training part, which consisted of an explanation and then four blocks (~20 min), then continued with eight more passive chroma, height, and control blocks (~40 min), and finally had two blocks of the attend chroma part, following another brief explanation (~10 min).

Behavioral Analysis

Behavioral responses were collected, and the individual d' sensitivity index was calculated both for the ear training and attend chroma parts. A correct detection of G was

considered as a hit, and detecting G when the note was C was a false alarm (FA). To avoid infinite d' values, a maximal performance of $p\text{Hits} = 1$ was replaced by $1 - (0.5 \text{ mistakes} / \text{the total number of Gs})$ and a $p\text{FA} = 0$ was replaced by $0.5 \text{ mistakes} / \text{the total number of Cs}$. As a result, the maximal d' (no misses and no false alarms) was 5.136 for the ear training part and 5.26 for the attend chroma part. The correlation between d' s of the ear training and of the attend chroma tests was calculated using Spearman's rank correlation. Post hoc, participants whose d' in the attend chroma part was higher than the average were considered as good performers. In addition to the planned analysis of all participants, we also report the exploratory results restricted to the good performers.

EEG Recording and Preprocessing

EEG was recorded from 64 preamplified Ag/AgCl electrodes using an Active 2 system (BioSemi, the Netherlands), mounted on an elastic cap according to the extended 10–20 system, with the addition of two electrodes over the mastoids and a nose electrode. Five additional electrodes tracking electrooculographic activity were placed on the outer canthi of the right and left eyes, below the center of both eyes, and above the center of the right eye. The EEG signal was sampled at a rate of 512 Hz (24 bits/channel), with an online antialiasing low-pass filter set at one fifth of the sampling rate, and stored for offline analysis.

EEG preprocessing was conducted using BrainVision Analyzer 2.0 (Brain Products) and MATLAB (2016b, MathWorks). First, detrending was applied using MATLAB, subtracting long-term linear trends from each block, thus zeroing its edges and avoiding discontinuities after concatenation. Then, further preprocessing was done in Analyzer, using the following pipeline: 0.1 Hz high-pass, zero-phase-shift second-order Butterworth filter; referencing to the nose electrode; correction of ocular artifacts using independent component analysis (Jung et al., 2000) based on typical scalp topography and time course; and discarding epochs that contained other artifacts (rejection criteria: absolute difference between samples $> 100 \mu\text{V}$ within segments of 100 msec; gradient $> 50 \mu\text{V}/\text{msec}$; absolute amplitude $> 120 \mu\text{V}$; absolute amplitude $< 0.5 \mu\text{V}$). Finally, using MATLAB, a 1–20 Hz band-pass zero-phase-shift fourth-order Butterworth filter, optimal for MMN analysis (e.g., as in Deouell, Parnes, Pickard, & Knight, 2006), was applied to the continuous data, followed by segmentation and averaging.

EEG Analysis

We calculated ERPs locked to auditory stimulus presentation. Data were parsed into segments beginning 100 msec before the onset of tone presentation, which served as baseline. In the passive MMN part, segments were

500-msec long, including baseline, and difference waves were calculated by subtracting the control from the deviant waveform. After rejecting segments that included artifacts, the average number of segments per participant was 269 for deviant and control in the chroma deviation condition, 261 for deviant and control, and 1045 for standards in the height deviation condition. In the attend chroma part, segments were 1100-msec long, including baseline, with an average of 54 segments per subject for targets (G) and 215 for nontargets (all four types: C4–C7 together) after artifact rejection. The ERPs of the attend chroma part were computed using only correct responses (hits—targets that the participant detected by a button press—and correct rejections—nontargets for which there was no response). This resulted in 53 segments for targets and 203 for nontargets on average per participant.

Statistical analysis of the MMN waveforms was performed on electrode Fz, as accepted in the literature (e.g., Näätänen, Paavilainen, Rinne, & Alho, 2007). Difference waves were calculated, subtracting from the deviant in each condition its matched control. For statistical assessment, a *t* sum cluster-based permutation test (Maris & Oostenveld, 2007) was run on the difference wave epoch of 50–300 msec from stimulus onset, using 10,000 permutations and a probability threshold of .01 for inclusion in a cluster. If a significant cluster was found ($p < .05$), a scalp topography was calculated, averaging over the temporal extent of the cluster. For assessing the P3b component in the attend chroma part, a similar procedure was used, subtracting the average waveform of the nontargets from the target waveform, but the permutation test was run on an epoch of 200–1000 msec from stimulus onset.

Results

Musicians Are Able to Perceive Pitch Chroma of Pure Tones

Results from the ear training part of Experiment 1 indicate that participants were able to classify pure tones from four octaves to either C or G well (average $d' = 3.36$, $SD = 1.64$). Not surprisingly, the five AP listeners did the task near perfectly ($d' = 4.9$, $SD = 0.28$; Figure 2). Results from the attend chroma part confirm that participants could detect the target G5 among the four nontarget Cs (C4, C5, C6, and C7), in a similar setting to that of the chroma deviation condition in the passive MMN part (mean $d' = 3.87$, $SD = 1.46$ for all participants; $d' = 5.06$, $SD = 0.32$ for AP listeners; Figure 2). The performance in the two tests was highly correlated (Spearman's $\rho = .82$, $p = 6.7 \times 10^{-6}$; Figure 2).

Pitch Chroma Deviations Do Not Induce an MMN

In the passive MMN part of Experiment 1, the difference waveform of electrode Fz, subtracting the response to the G tone in the control block from the identical G tone, which served as a deviant in the chroma deviation block, did not show any significant negative deflection (Figure 3B). No significant cluster was found in the *t* sum cluster-based permutation test, even when the analysis was restricted to only good performers, whose d' was larger than average ($n = 13$, all d' s > 3.87 , mean $d' = 4.85$, $SD = 0.4$, see Figure 3B).

In contrast to chroma deviations, a significant MMN was measured for height deviations, as expected (Figure 3A). The difference waveform of electrode Fz, subtracting the response to the A tone in the control

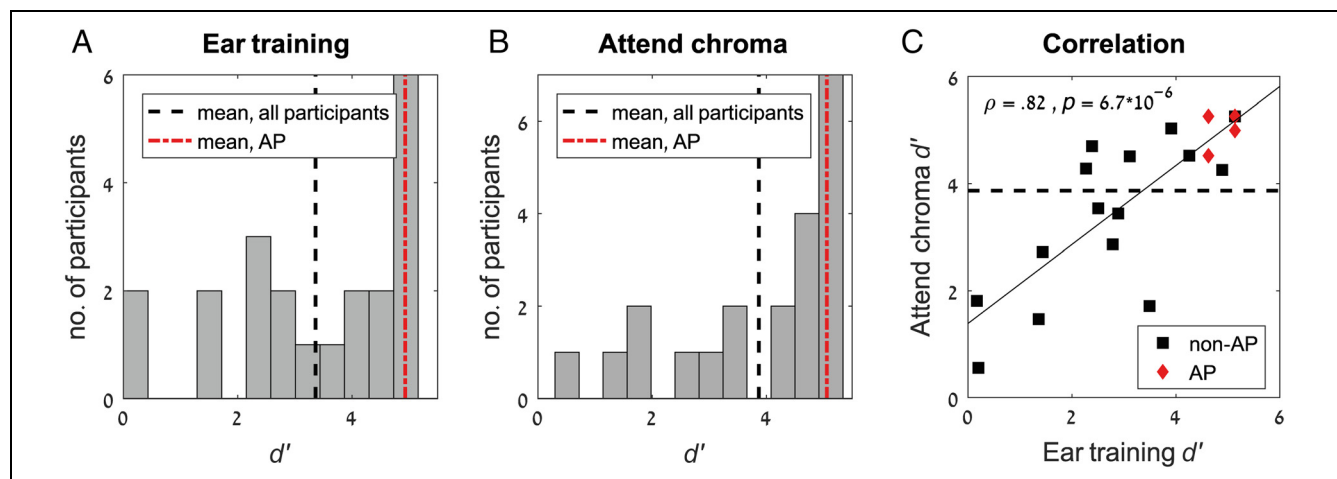


Figure 2. Behavioral results, Experiment 1. (A) Histogram of d' s in the ear training part. The black dashed line is the mean d' of all 21 participants. The red dashed line is the mean d' of the five AP listeners. (B) Histogram of d' s in the attend chroma part. Dashed lines are the same as in A. (C) Correlation between the d' s of the ear training and attend chroma parts. Each dot represents one participant. The black squares are non-AP listeners, and red diamonds are AP listeners. Four participants are overlapping at the point of maximal performance—two AP and two non-AP listeners. Spearman's rho and the p value of the correlation are denoted. The black dashed line is the mean d' in the attend chroma block of all participants, above which participants were defined as “good performers” for later EEG analysis.

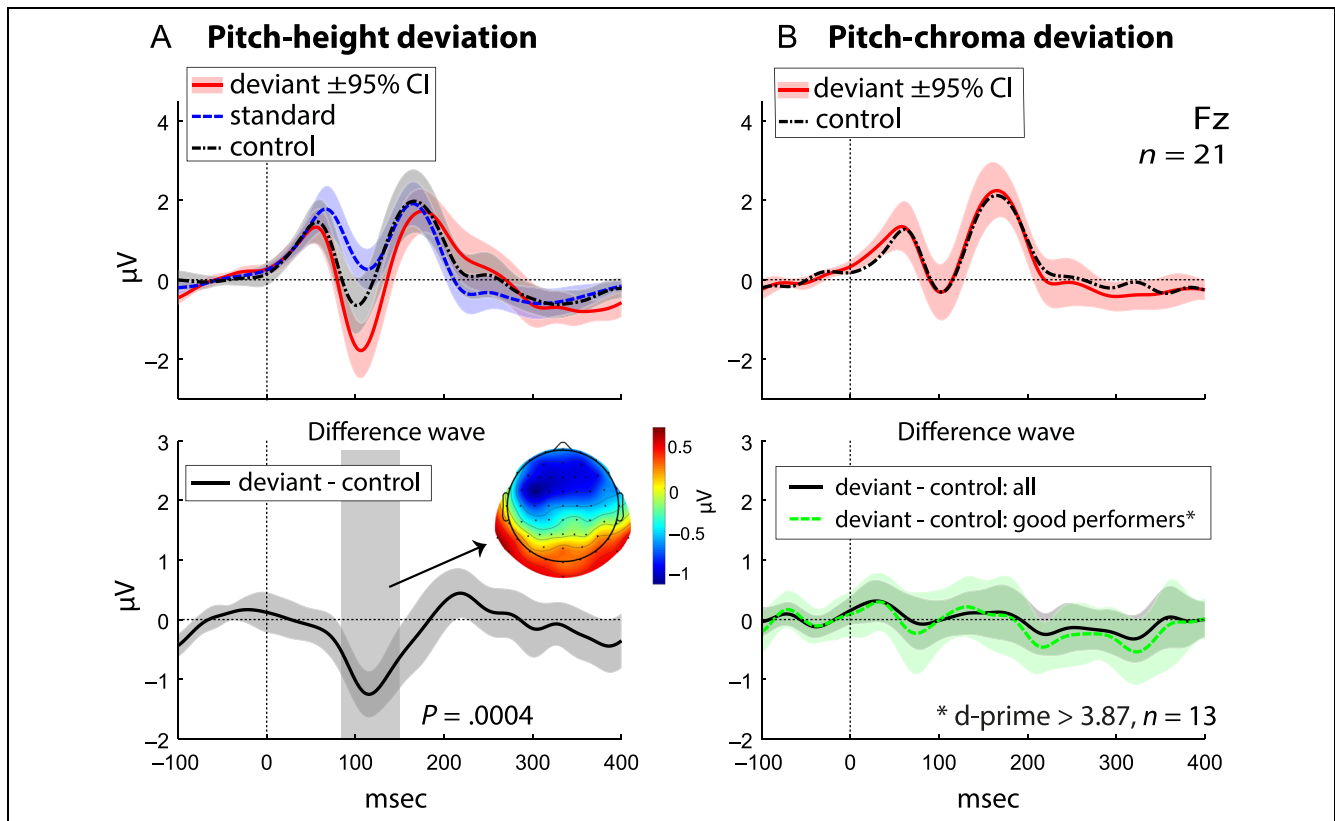


Figure 3. Passive MMN part of Experiment 1. (A) Height deviation elicits an MMN. Top: ERPs of the deviant (A5, 880 Hz, comprising 20% of the tones in the sequence), standard (D5, 587.3 Hz, 80%), and control (20%, same note as the deviant, see Methods) are presented. ERPs are grand averages of 21 participants, electrode Fz. The shaded area is a confidence interval of 95% around the mean. Bottom: Difference wave ERPs subtracting control from the deviant. The shaded rectangle marks an epoch of a significant cluster of t sum, according to a permutation test (see Methods, as in Maris & Oostenveld, 2007). The topography on the right top corner of the graph is that of the average voltage within the significant cluster epoch. (B) Chroma deviation does not elicit an MMN. Top: Details are similar to A. Bottom: Details are similar to A. In dashed green, the difference wave of only good performers (defined as having d' in the attend chroma part greater than the mean, 13 participants) is plotted on top of the difference wave of all participants.

block from the identical A tone, which served as a deviant in the height deviation block, showed a significant ($p = .0004$) negative deflection between 83 and 150 msec peaking around 115 msec, with an amplitude of $-1.25 \mu\text{V}$. The topography of this response was typical to the MMN (Figure 3A), with a frontal negativity that flips in the mastoid channels.

The ERPs in the attend chroma part showed typical N2–P3b responses with a parietal maximum (Figure 4A) in responses to the target Gs compared with nontarget Cs.

In summary, in Experiment 1 we found no evidence for automatic detection of a chroma deviant G among the chroma standards C4–C7 despite the ability to detect it when it was task relevant, even in good performers who could detect the chroma deviant easily. This suggests that, unlike height, chroma is not processed automatically in an unattended stream.

EXPERIMENT 2

The chroma pair C and G is highly consonant. Thus, it might be the case that the chroma of C and G is too

similar and that less similar chroma pairs would be easier to discriminate in the unattended stream. To exclude this possibility and generalize our results, we ran another similar experiment using the chroma pair C and F#. This pair constructs the “tritone” interval—considered as the most dissonant interval.

Methods

The design and rationale of this experiment followed those of Experiment 1. The main difference was the replacement of the fifth interval chroma pair (C–G) with a tritone interval chroma pair (C–F#). However, to increase the efficiency of the experiment as well as to increase the generalizability of the results, we introduced a few additional minor modifications, which we did not expect to influence the results.

Participants

Twenty-eight healthy musicians participated in the Experiment 2 and were paid 40 shekels ($\sim\text{US}\$12$) per

hour. Participants were recruited in a similar way to Experiment 1 and with the same inclusion criteria, except that AP was an exclusion criterion (to increase sample homogeneity). One participant was excluded because of excessive artifacts and falling asleep during the experiment. The analysis included the data of 27 participants (15 women, mean age = 25 years, $SD = 2$ years). All participants self-reported normal hearing and no history of neurological disorders. The experiment was approved by the ethical committee of the faculty of social science

at the Hebrew University of Jerusalem, and informed consent was obtained after the experimental procedure was explained.

Stimuli and Apparatus

The experimental setting was similar to that of Experiment 1. Experiment 2 included both pure tones, with the same characteristics as in Experiment 1 and complex tones. The complex tones consisted each of five frequencies,

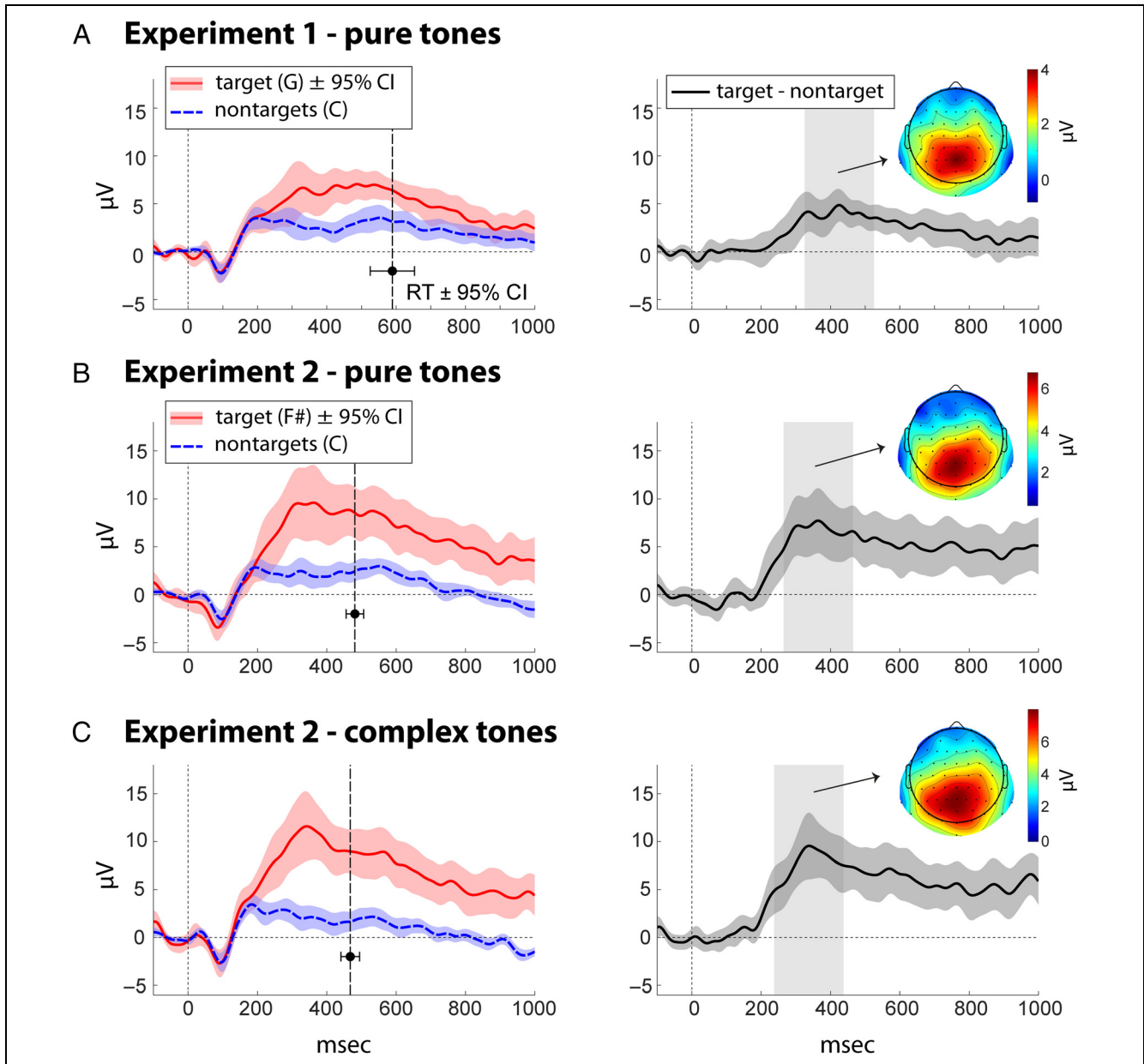


Figure 4. Attend chroma part, Experiments 1 and 2. (A) Experiment 1. Left: ERPs of the target (G5, 784 Hz, comprising 20% of the tones in the sequence) and nontargets (C4–C7, 261.6, 523.2, 1046.5, and 2093 Hz, 80%) are presented. ERPs are grand averages of 21 participants, electrode Pz. All ERPs are calculated including only correct responses both for deviants (hits) and for standards (correct rejections). The shaded area around the ERPs depict the 95% confidence interval around the mean voltage. The vertical dashed line signifies the mean RT for hits, and the small horizontal line crossing it signifies the 95% confidence interval around the mean. Right: Difference wave ERPs subtracting nontargets from the target. Epochs for topography plotting (shaded gray areas plots) were selected as a range of 200 msec around the peak. (B) Experiment 2—pure tones. Left: Details are similar to A except that the target was F#5, 740 Hz. Right: Details are similar to A. (C) Experiment 2—complex tones. Details are similar to B.

spaced an octave apart from each other. These frequencies were synthesized under a Gaussian spectral envelope (with a logarithmic frequency axis), such that the middle frequency had the strongest power, the neighboring two octaves (higher and lower) were lower by 6.78 dB, and the two next ones were lower by 27.15 dB comparing to the middle tone. These stimuli are sometimes known as “Shepard tones.” They were originally designed by Shepard (1964) to induce the illusion of increasing pitch under a constant spectral envelope. However, we shifted the Gaussian center frequencies of the tones such that each center was placed over the frequency of the matched pure tone condition (see Figure 1). This timbre allowed us to construct complex tones using frequency components of only the same chroma and avoid the possibility of overlapping harmonics between tones of different chromas. All other parameters and procedures were similar to Experiment 1.

Experiment Design

Ear training. In this part, similar to Experiment 1, participants had to classify tones according to their pitch chroma. Eight tones spanning four octaves were presented; four Cs (similar to Experiment 1) and four F#s (370, 740, 1480, and 2960 Hz, corresponding to F#4, F#5, F#6, and F#7, respectively), replacing the Gs in Experiment 1. All details were similar to Experiment 1, except for the following: Participants responded by pressing one of two buttons assigned to either C or F#. Participants placed two fingers of their dominant hand on two neighboring keyboard keys—the index finger was assigned to chroma C and the middle finger was assigned to chroma F#. After each response, the correct answer—C or F#—replaced the fixation cross for 800 msec. Green letters were used for a correct answer, red for a wrong answer, and black if the participant did not respond within the maximal allowed time—3 sec. There were four blocks of pure tones and four blocks of complex tones (Figure 1, see Stimuli and Apparatus section for a detailed description of the complex tones), 50 tones in each block. The order of the blocks was counterbalanced between participants, such that for half the order it was ABABBAAB and for the other half it was BABAABBA (where A and B denote pure and complex tone blocks, respectively).

Passive MMN. This part was similar to the passive MMN part of Experiment 1, except for the following details. There were six pure tone blocks and six complex tone blocks. For each type of tones, the six blocks consisted of two height deviation blocks, two chroma deviation blocks, and two control blocks. The pitch height deviation blocks contained 80% standard tone B4 (493.8 Hz) and 20% deviant tone D6 (1174.6 Hz). The pitch chroma deviation blocks included five tones—four standard tones having chroma C, from four octaves (same as in

Experiment 1), and the deviant tone F#5 (740 Hz). The five tones appeared each 20% of the times. The control blocks included five tones; Db4, C5, F#5, D6, and B6 (277.9, 523.2, 740, 1174.7, and 1975.6 Hz, respectively) each presented 20% of the times. The F#5 tone served as the control tone for the chroma deviant, and the D6 served as the control for the height deviant. In the case of complex tone blocks, the same frequencies listed above were the central and highest level component, accompanied each by four other components, one and two octaves above and below, with lower levels (see Stimuli and Apparatus section and Figure 1). Each block included 550 trials presented with an SOA of 400 msec. Because each block was presented twice, this resulted in 220 trials for each deviant or its comparable control. Each block lasted 220 sec, and there was 30 sec of rest between blocks (or longer at the participant discretion). The order of the blocks was counterbalanced between participants, such that for half the order it was ABCDEFABCDEF and for the other half it was FEDCB AFEDCBA, where A, B, and C stand for height deviation, chroma deviation, and the control block, respectively, all with pure tones, and D, E, and F stand for height deviation, chroma deviation, and the control block, respectively, all with complex tones.

Attend chroma. This part was similar to the attend chroma part of Experiment 1, except for the following: F#5 replaced G5, the SOA was 1000 msec, and there were four blocks—two with pure tones and two with complex tones—75 trials each. This resulted in a total of 30 pure tone targets and 30 complex tone targets. The order of block presentation was counterbalanced between participants such that for half it was ABAB and for the other half it was BABA (where A and B stand for pure and complex tone blocks, respectively).

To summarize, participants started with the ear training part, which consisted of explanation and then eight blocks intermixed between pure and complex tones (~20 min), then continued with the 12 passive MMN blocks while viewing a silent film (~50 min). Finally, four blocks of the attend chroma part followed another brief explanation (~10 min).

Behavioral Analysis

Behavioral responses were analyzed similarly to Experiment 1 (Behavioral Analysis section), except that in Experiment 2, the post hoc selection of “good performers” was based on the average d' over pure and complex tones in the attend chroma part. In addition, in Experiment 2, performance in the ear training and attend chroma parts was compared between pure and complex tones. d' s were calculated separately for pure and complex tones and statistically compared using a paired-samples sign test.

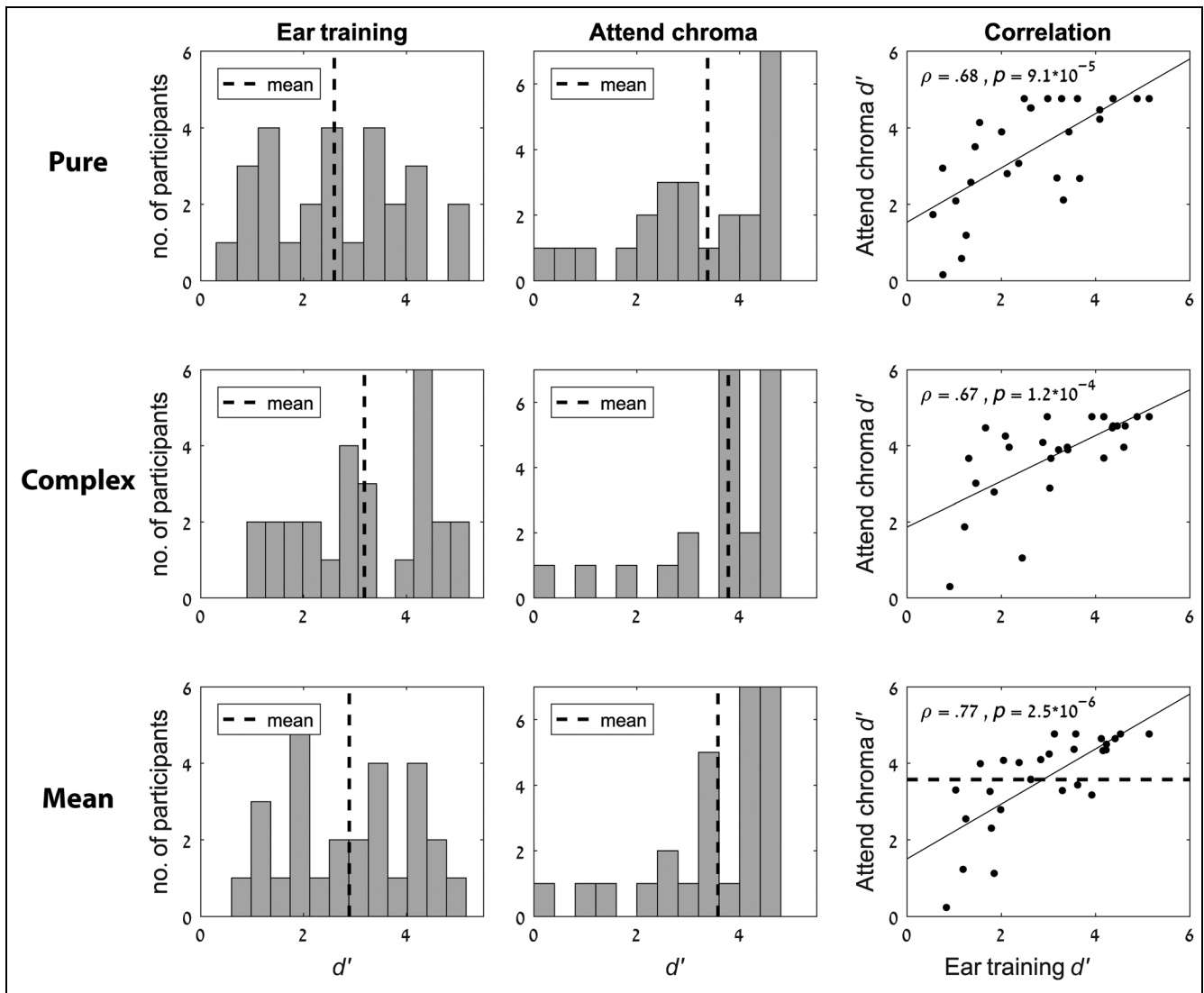


Figure 5. Behavioral results, Experiment 2. Top row: Pure tones performance. Middle row: Complex tones performance. Bottom row: Mean performance of pure and complex tones per each participant. Left column: Histograms of d' 's in ear training part. Black dashed line is the mean d' of all 27 participants. Middle column: Histograms of d' 's in the attend chroma part. Dashed line similar to A. Right column: Correlation between d' 's in the ear training and attend chroma parts. Each dot represents one participant. Spearman's rho and the p value of the correlation are denoted. The dashed black line in the lower correlation plot is the mean d' across participants of mean pure and complex tone performance per participant, in the attend chroma block. Participants above this line were defined as “good performers” for later EEG analysis.

EEG Recording and Preprocessing

EEG recording and preprocessing were identical to Experiment 1 (EEG Recording and Preprocessing section).

EEG Analysis

EEG analysis was similar to Experiment 1 (EEG Analysis section). In the passive MMN part, after artifact rejection, the average number of segments per participant, for pure tones, was 217 for deviant and 218 for control in the chroma deviation condition, 218 for deviant or control, and 874 standards in the height deviation condition. For complex tones, it was 212 for deviant and 215 for

control in the chroma deviation condition, 216 for deviant, 214 for control, and 862 standards in the height deviation condition. In the attend chroma part, segments were 1100 msec long, including baseline, and an average number of 27 segments per participant for target (F#) and 108 for nontargets (all four types: C4–C7 together) remained after artifact rejection for pure tones or complex tones. We then calculated ERPs in the attend chroma part using only correct responses (hits—targets that the participant detected by a button press—and correct rejections—nontargets for which there was no response) and the remaining average number of segments per participant was 23 targets and 107 nontargets for pure tones and 26 targets and 106 nontargets for complex tones.

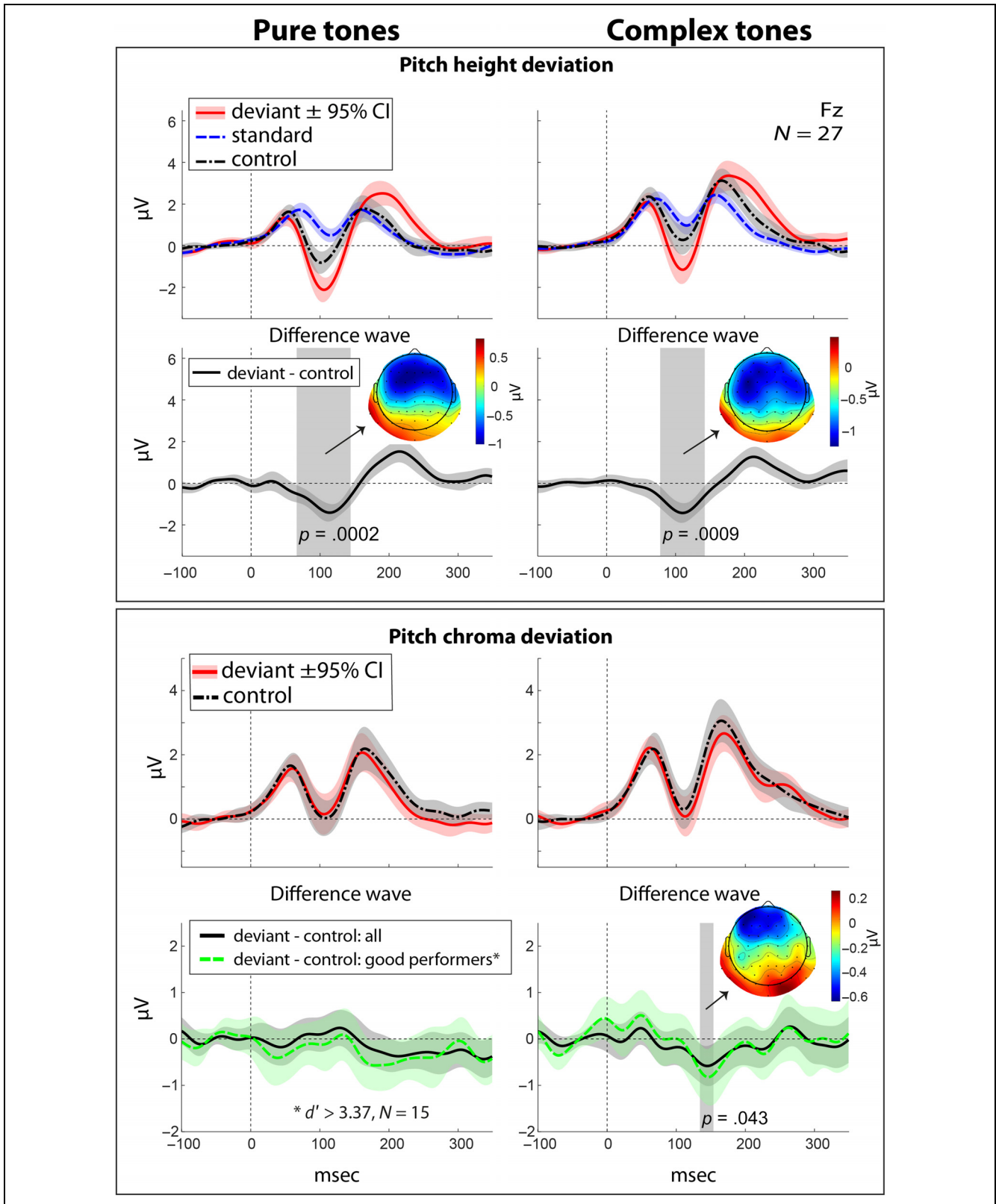


Figure 6. Passive MMN part of Experiment 2. Left column: Pure tones. Right column: Complex tones. Top rectangle: Height deviation. ERPs of the deviant, standard and control (see Methods for technical details). ERPs are grand averages of 27 participants, electrode Fz. The shaded area is a confidence interval of 95% around the mean. The shaded rectangle marks an epoch of a significant cluster of t sum, according to a permutation test (see Methods, as in Maris & Oostenveld, 2007). The topographies are of the average voltage within the significant cluster epoch. Bottom rectangle: Chroma deviation. Details are similar to the latter. In dashed green, the difference wave of only good performers (defined as having d' in the attend chroma part greater than the mean, 16 participants) is plotted on top of the difference wave of all participants. Clusters analysis was run on both difference waves (see text), but details here (p value, significant cluster epoch and topography) refer to the analysis of all participants together.

Results

Pure Tones—Behavior

Results from the ear training part of Experiment 2 indicate that musicians were able to classify pure tones from four octaves to either C or F# (mean d' = 2.6, SD = 1.32; these d' differed from 0 significantly; p = 5.6×10^{-6} , Wilcoxon signed-rank test). Results from the attend chroma part again confirmed that the participants could detect the target F#5 among the four nontarget Cs, in a similar setting to that of the chroma deviation block in the passive MMN part (mean d' = 3.38, SD = 1.4). Spearman's correlation coefficient between these two tests across participants was .68 (p = 9.1×10^{-5} ; Figure 5, first row).

Pure Tones—No Chroma MMN for the Tritone Interval

In the passive MMN part of Experiment 2, similar to Experiment 1, the difference waveform of electrode Fz, subtracting the response to the F# tone in the control block from the identical F# tone, which served as a deviant in the chroma deviation pure tone block, did not show any significant negative deflection (Figure 6, bottom rectangle, left), even when the analysis was restricted to only good performers, whose d' s in the attend chroma part were larger than average (n = 15, d' > 3.37, mean d' = 4.44, SD = 0.41; Figure 6). Some late negative trend around 200 msec was observed in the difference waveforms (Figure 6, bottom rectangle, bottom left), but this trend was not significant.

In contrast to chroma deviations and similar to Experiment 1, a significant MMN was measured for height deviations using pure tones (Figure 6, top rectangle, left). The difference wave of electrode Fz, subtracting the response to the D tone in the control block from the identical D tone, which served as a deviant in the height deviation block, showed a significant (p = .0002) negative deflection between 78 and 142 msec peaking around 115 msec, with an amplitude of $-1.4 \mu\text{V}$. The topography of this response was typical to the MMN (Figure 6, top rectangle, left) with a frontal negativity that flips in the mastoid channels.

The ERPs in the attend chroma part showed typical N2–P3b responses with a parietal maximum (Figure 4B) in responses to the targets compared with nontargets.

In summary, in Experiment 2, we replicated the result of Experiment 1, in which no MMN was elicited for a chroma deviation using pure tones and generalized it to the dissonant chroma pair C and F#.

Complex Tones Improve Performance

In the ear training part of Experiment 2, performance was somewhat better for complex tones (mean d' = 3.18, SD = 1.26) than for pure tones (mean d' = 2.6, SD = 1.32). The difference was significant (paired-sample

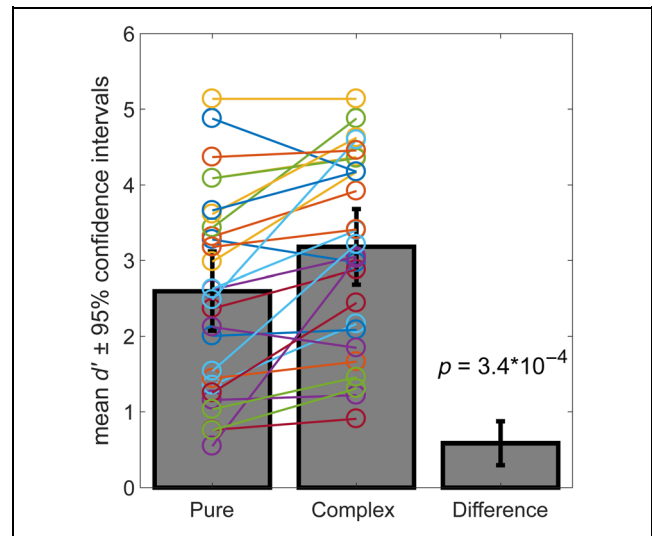


Figure 7. Comparing ear training performance of pure versus complex tones. d' s of performance in the ear training part of Experiment 2, discriminating notes from four octaves to either C or F#. Mean d' s in blocks using only pure or complex tones and the mean of differences (d' complex – d' pure, per individual) are presented in the bar graph. Error bars represent confidence intervals of 95% around the mean. The colorful lines connecting the circles over the bar graph represent all individual participants. The p value is of a paired-samples sign test.

Wilcoxon signed-rank test, p = 3.4×10^{-4}), with 23 of 27 participants showing improved performance for complex tones (Figure 7). In the attend chroma part, the average d' was also higher for complex tones (d' = 3.78, SD = 1.15) than for pure tones (d' = 3.37, SD = 1.4), and this difference was also significant (paired-sample Wilcoxon signed-rank test, p = .017).

Do Complex Tones Elicit a Small Chroma MMN?

Using complex tones, in the passive MMN part of Experiment 2, a small, marginally significant (p = .043) negativity was found in the difference waveform of chroma deviations between 134 and 154 msec peaking at 144 msec with a peak amplitude of $-0.58 \mu\text{V}$. This was in contrast to the same condition using pure tones, for which we did not get a significant MMN (Pure Tones—No Chroma MMN for the Tritone Interval section). The topography of the significant cluster was consistent with a typical MMN topography but was more frontal and localized than that of the height MMN (Figure 6, bottom rectangle, right). Restricting the analysis to “good performers” (n = 15), having an above average d' in the attend chroma block (see Methods section), a slightly larger negativity was found at similar latencies, peaking at 146 msec with a larger absolute amplitude of $-0.82 \mu\text{V}$ (Figure 6, bottom rectangle, right, green trace). Significance was not tested for the smaller number of participants.

A significant MMN was obtained for height deviations using complex tones (Figure 6, top rectangle, right). The

difference wave of electrode Fz, subtracting the response to the D tone in the control block from the identical D tone, which served as a deviant in the height deviation block, showed a significant ($p = .0009$) negative deflection between 66 and 144 msec peaking around 110 msec with a peak amplitude of $-1.43 \mu\text{V}$.

The ERPs in the attend chroma condition using complex tones showed typical N2–P3b responses with a parietal maximum (Figure 4C) in responses to the targets compared with nontargets, with a similar pattern for the pure and complex tones.

DISCUSSION

We studied the automaticity of chroma processing in the human brain, using the MMN as a signature for automatic, nonintentional (or preattentive) processing. In two experiments, we found that trained musicians were able to discriminate the chroma of pure tones spread across four octaves. However, despite the ability to discriminate pure tones based on chroma, we found no neural evidence for automatic detection of the pure tone chroma deviants, even for higher-than-average performers and even when the deviant was musically dissonant comparing to the standard. Thus, we find no evidence that chroma is a dimension, which is processed automatically, in unattended streams (at least as indexed by the MMN).

The MMN as a Proxy for Preattentive Processing

The MMN is commonly used to tap for automatic processing, usually using auditory stimuli. The word “automatic” is used here to signify processes that take place regardless of the task and do not require attention. Typically, MMN studies use an oddball paradigm, in which some rule of regularity is established during the sequence, and some rare deviant stimuli violate this rule. These paradigms are passive—the participant is instructed to ignore the stimuli and perform a different task (such as viewing a silent film as in our case). If the rare change in the stimulus dimension that established the regularity elicits the MMN during passive listening, then this is an indication for automatic processing of that dimension.

In general, it is accepted that any discriminable change will elicit the MMN (Näätänen et al., 2007). For an example, it was shown that the minimal sound frequency difference that elicits an MMN correlates with perceptual limits (Näätänen et al., 2007; Sams, Paavilainen, Alho, & Näätänen, 1985). MMN was shown for almost any physical auditory feature, for example, intensity, duration (Näätänen et al., 2007), frequency (Sams et al., 1985), and spatial location (Deouell et al., 2006; Schröger & Wolff, 1996). Beyond simple physical features, several studies show MMN for more abstract regularities. For example, a locally ascending note among a descending

sequence of notes (Tervaniemi, Maury, & Näätänen, 1994). Thus, automatic processing of musical features was suggested, such as melody contours (Tervaniemi, Rytönen, Schröger, Ilmoniemi, & Näätänen, 2001) and even music syntax (Koelsch, 2009; Poulin-Charronnat et al., 2006). This “musical MMN” was shown to be enhanced both by perceptual learning in short-term training and by long-term expertise (Tervaniemi et al., 2001).

In the MMN literature, there are very few examples of auditory features that do not elicit the MMN. For instance, spectral modulations along a 1-sec-long tone induced an MMN only if occurring within 400 msec after sound onset. Otherwise, no MMN was measured in the absence of attention (Grimm & Schröger, 2005). However, the literature is missing a detailed characterization of the limits of automatic processing. We found here that a “perceivable” change in chroma does not elicit the MMN. The discrepancy between the overt identification of the deviants while task relevant and the lack of MMN in the unattended condition may indicate that grouping pure tones according to their chroma is a task that involves higher cognitive processes, such as attention, working memory, and acquired associations. Future studies exploring the general limitations of the MMN system might elucidate the processes underlying pitch chroma processing.

Comparison with Previous Studies of Pitch Chroma

Pitch chroma expresses the property of octave equivalence. The octave interval serves as a basic structure in almost any modern music system (Wallin, Merker, & Brown, 2000). Yet, it is not clear whether the perception of octave equivalence is biologically innate. Behavioral evidence for chroma processing is mixed. On the one hand, octave generalization of pure tones was shown in humans—musicians and nonmusicians (Hoeschele et al., 2012), and even infants (Demany & Armand, 1984). On the other hand, 4- to 9-year-old children rated tone similarity due to height proximity with no evidence for octave equivalence (Sergeant, 1983). Other similarity rating studies gave evidence for octave equivalence perception in trained musicians (Allen, 1967) but failed to show robust results in nonmusicians (Kallman, 1982; Krumhansl & Shepard, 1979; Allen, 1967). Octave generalization was sparsely shown in other mammals: Monkeys (Wright et al., 2000) and rats (Blackwell & Schlosberg, 1943) showed evidence for generalization, but avians like chickadees (Hoeschele, Weisman, Guillette, Hahn, & Sturdy, 2013) and European starlings (Cynx, 1993) did not. Thus, it is still an ongoing debate whether octave equivalence is a general perceptual property, dependent on physiological constraints, or is a higher level concept dependent on learning, exposure, and other cognitive and cultural factors (Sergeant, 1983).

The helix model, discussed in the Introduction, implies a contribution of both height and chroma to the neural

representation of pitch. Nevertheless, studies of the neural organization underlying pitch mostly concentrate on height. Because pitch, although related to frequency content, is not indexed by the tonotopic organization of the early auditory system, various attempts have been made to find a periodotopic organization (for an exhaustive review, see Schnupp, Nelken, & King, 2011, chap. 3). Such a topographic representation of sound periodicity, the best correlate of pitch perception, is usually thought of as a monotonous gradient from low to high fundamental frequencies and thus represents height.

In contrast, the neural underpinnings of chroma are largely unknown. A neural structure encoding pitch chroma is expected to generalize across octaves, that is, show a similar firing pattern for sounds spaced an octave apart, independent of other auditory parameters, such as timbre or height. It was anecdotally suggested that a structure in the ventral nucleus of the lateral lemniscus found in gerbils has a helical anatomical structure corresponding to the pitch helix (Langner & Ochse, 2006). To our knowledge, no such neural correlate of pitch chroma was found in humans, but several recent imaging studies suggested cortical representation of chroma.

A recent fMRI study found clusters of voxels tuned to pairs of frequencies an octave apart, spread all over the supratemporal plane (Moerel et al., 2015). The authors hypothesized that multiplex spectrally tuned neuronal populations (Moerel et al., 2013) in these voxels contribute to the percept of octave equivalence. Such populations of neurons could have been an appealing mechanism for generalizing across octaves and detecting chromatic regularity, allowing for a mismatch to be detected. However, in addition to octave tuned voxels, clusters of voxels tuned to other intervals were observed as well, both with or without harmonic relations. The amount of octave tuned voxels did not exceed the amount of the voxels tuned to other intervals. Therefore, the results of Moerel et al. (2013, 2015) do not give a special status to the octave interval relative to other intervals.

Warren et al. (2003) suggested, using an fMRI adaptation paradigm, that chroma is represented anterior to primary auditory cortex, whereas height is represented posterior to it. However, it is not clear whether the regions of activation associated with chroma in their study represent chroma per se. This ambiguity stems from the fact that chroma was manipulated by inducing small alterations of the fundamental frequency within one octave and, as a result, was not independent of height. Furthermore, considering the poor temporal resolution of fMRI, it is not clear whether the reported activity represents early and automatic, or late processing that depends on attention.

Using EEG, Briley et al. (2013) found chroma-based adaptation of the N1–P2 components of the auditory evoked potentials (~100–200 msec poststimulus), yet only for complex tones and not for pure tones. In our study as well, a small but significant MMN was measured for

chroma deviations of complex tones, but not pure tones (Figure 6). Behavioral results from the ear training task of our Experiment 2 indicate also that chroma of complex tones is slightly easier to perceive than that of pure tones (Figure 7). These results require to spell out explicitly the relationships between pure tones, complex tones, and chroma.

Briley et al. (2013) argue that the adaptation effect they found when using complex tones was driven by chroma-sensitive neurons. Because, in their view, pitch is inextricably related to timbre (spectral content), these neurons did not respond to pure tones. We argue that genuine chroma-selective neurons should generalize over timbre and therefore should show octave equivalence regardless of spectral content. Indeed, chroma can be overtly and accurately perceived with pure tones, as reported in our study. Thus, we believe that the neuronal resources that were adapted in the experiments of Briley et al. (2013) cannot be truly chroma-sensitive neurons.

Instead, we maintain that measuring chroma-based adaptation using complex tones mixes up chroma-selective neural representation with adaptation based on physical similarity: frequency overlap between the partials that compose the complex tones or the temporal structure of the resulting spike trains. The fundamental frequencies of two tones having the same chroma in consecutive octaves have the ratio of 1:2. Therefore, all of the harmonics of the higher tone, including the fundamental frequency, are contained among the harmonics of the lower tone. For this reason, two tones having the same chroma may share neural representations just because of physical similarity on the frequency dimension.

In the current study, we used the MMN paradigm, which allowed us to concentrate on regularity extraction rather than adaptation. We overcome the confound of adaptation by comparing the deviant in the experimental blocks to a control that undergoes a comparable amount of adaptation to that of the deviant, instead of comparing the deviant to the standards, which may be substantially more adapted due to physical similarity in both the height condition and complex tones chroma condition. The small chroma MMN found in the complex tones condition could result from height regularity extraction in the frequency bands corresponding to the common harmonic components of the two sounds and thus does not necessarily imply preattentive chroma regularity extraction. Specifically, the complex tones we used had the so called “Shepard tone” timbre (Shepard, 1964); each tone was composed of five frequency components in octave relationships, from five consecutive octaves, under a Gaussian spectral envelope. The center frequencies of the Gaussian envelopes were located at the pure tone frequencies used in the pure tone condition. Figure 8 shows how in this regime all standards share a component at one of the central C frequencies, creating a simple height regularity at this frequency, which is violated by the F# tones. In consequence, frequency-specific

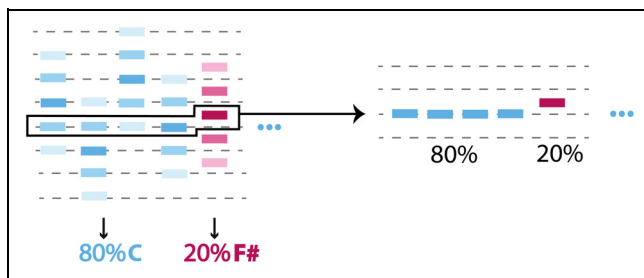


Figure 8. Chroma MMN in complex tones can be driven by height MMN mechanisms. As in Figure 1, the y-axis is a log frequency axis, and the x-axis represents time. Dashed lines represent the C octaves. Colored lines represent the frequency components of the tones, where the color represents chroma and the brightness of the color represents the intensity of each frequency component. In the left part of the figure, a complex tone block in the passive MMN part is illustrated. The right part demonstrates that, considering a narrow frequency band in the complex stimuli, a scenario similar to the height MMN protocol occurs within the complex stimuli. Therefore, an apparent “chroma” MMN in the complex tones does not imply chroma-selective neurons but can be derived by simple frequency-selective neurons.

neurons provide a representation that is sufficient to detect this rule violation and no chroma-specific neurons are required.

Is Pure Chroma Perception Dependent on Attention?

An important feature of our study relative to all of the above findings of apparent chroma-related neuronal tuning is that the participants’ attention was directed to a primary visual task. In contrast, the previous studies used an active listening task (Moerel et al., 2015; 1-back task) or did not use any task at all (Briley et al., 2013; Warren et al., 2003), and therefore, attention was probably directed toward the stimuli. The distinction of automatic from attention-dependent representations is important because it probes the level of processing. Automatic processes can largely be considered “bottom-up” in contrast to task-dependent top-down effects. The lack of evidence for automatic processing of chroma in contrast to height indicates that chroma and height have fundamentally different neural representations, probably located at different stages of the processing hierarchy. We suggest that chroma is a higher-level percept dependent on human cognitive factors such as attention.

It might be the case that octave equivalence is a cognitive concept that develops due to the low-level physical similarities between complex tones with the same chroma. As discussed above, the spectral content of harmonic tones, and hence of most natural tones having the same chroma, overlaps considerably. These physical similarities give rise to automatic processing, which might facilitate behavioral detection of chroma (Figure 7). Consequently, the concept of chroma emerges and can

then be transferred and generalized to all pitch-evoking stimuli, including pure tones, yet this requires higher-level, nonautomatic, cognitive processes.

Limitations and Future Work

One of the limitations of this study is that our main result—the absence of an MMN—is a null result and therefore cannot be easily interpreted as strong evidence against automatic processing of chroma. Although caution must be exerted when interpreting null results, we note that these results were replicated in two separate groups and that they were obtained in well-trained musicians, who were further trained during the experiment and selected for being able to discriminate the deviance with high accuracy. In consequence, we failed to find an MMN under the optimal conditions for its presence.

Recently, Bayesian statistics is becoming increasingly popular for using null results as evidence for the null hypothesis (e.g., Dienes, 2014). Our expected effect, the MMN, has a variable peak latency and amplitude, which depends on stimulus features. The temporal uncertainty requires therefore multiple comparisons to detect the MMN in novel conditions. The method we used for significance testing (Maris & Oostenveld, 2007) is commonly used in the EEG and MEG literature, because it accounts for the problems arising when analyzing continuous electrophysiological data, such as multiple comparisons over the sample points and uncertainty regarding the specific latency of effects. To our knowledge, there is not yet a standard method for calculating Bayes factors in scenarios in which both the latency and the effect size are unknown, and advances in this direction are needed. To convince ourselves in the reliability of our first set of results (Experiment 1), we replicated them. Indeed, Experiment 2 replicated the finding of no MMN to chroma of pure tones from Experiment 1 and included the condition of complex tones as a further control. The fact that we did measure a small but significant MMN to chroma of complex tones strengthens the validity of no MMN in the pure tone case, for the same participants.

It is of course possible that EEG is not sensitive enough to detect a weak mismatch response to the chroma deviations. In the future, automatic processing of chroma can be tested using ECoG—intracranial EEG—with the potential to observe more localized responses with a higher signal-to-noise ratio (e.g., Butler et al., 2011; Edwards, Soltani, Deouell, Berger, & Knight, 2005; Rosburg et al., 2005). Moreover, in a recent study, ECoG was used to separate the functionality of distinct cortical sources of the mismatch response, using the broadband high frequency signal, which is hard to detect on the scalp (Dürschmid et al., 2016).

In addition, in the average difference waveforms of the chroma MMN condition, a small trend of late negativity can be observed, starting around 200 msec and unfolding slowly until around 400 msec. This trend was more

prominent in Experiment 2 but did not reach significance in either of the experiments. These trends are not typical to the MMN effect—they are late, unfold slowly, and do not have a clear peak. It is possible though that they reflect some degree of processing of chroma in pure tones. Because they are later than a typical MMN, they might involve residual attention directed toward the stimuli. Otherwise, they could reflect preattentive processing that is late, small, and perhaps variable in latency between the participants. Future studies should examine whether these trends replicate and whether they depend on attention.

One potential concern in the interpretation of these data is the fact that the blocks aimed for studying chroma MMN differed from the blocks used for testing for height MMN in a number of ways. First, although the chroma condition isolated chroma from height, the height condition did not isolate height from chroma, as the deviant diverged from the standard in both height and chroma. This could cause a larger effect size of the height condition than of the chroma condition. To solve this, we could, in principle, run an experiment in which the deviant shares the chroma of the standard but is one octave higher (rather than a fifth or tritone, as used in Experiments 1 and 2). However, numerous studies have shown that increasing the frequency interval between standard and deviant results in an increased MMN (Tiitinen, May, Reinikainen, & Näätänen, 1994; Sams et al., 1985; see Loewy, Campbell, & Bastien, 1996, for an example of doubling the frequency), and thus, we can reliably expect the MMN in this case to be even larger than that found for the smaller frequency intervals tested here. In fact, the main reason for running the “height” condition was to verify that our participants showed the well-known MMN effect rather than to directly compare the height and chroma conditions. We indeed do not directly contrast them in any statistical analysis.

A second feature of the chroma condition that is different from that of the height condition was the nature of the standard in the chroma condition, which required generalization over a variation in height (for pure tones with the same chroma). Such generalization was not needed in the height condition, because the standard consisted always of the same physical stimulus. However, we believe that it is unlikely that the variability of the standards in the chroma condition can account for the absence of chroma MMN. Indeed, many previous studies showed that MMN can be obtained with variable standards (Daikhin & Ahissar, 2012; Pakarinen, Huotilainen, & Näätänen, 2010; Näätänen, Pakarinen, Rinne, & Takegata, 2004; Gomes, Ritter, & Vaughan, 1995; Winkler et al., 1990). In these studies, standards varied in a dimension orthogonal to the stimulus feature tested by the MMN (Pakarinen et al., 2010; Gomes et al., 1995; Winkler et al., 1990) or even in the tested feature itself (Daikhin & Ahissar, 2012; Winkler et al., 1990). In some studies, the standards varied in more than one feature,

for example, two features in Gomes et al. (1995), one of which was a frequency variability similar to our case. In the most extreme case, Pakarinen and colleagues (2010) designed a multifeature paradigm to test MMN elicited by eight auditory features within the same sound sequence, resulting in very large variability of the standards. Still, a robust MMN to all features was reported. Some studies (Daikhin & Ahissar, 2012; Winkler et al., 1990) did note decreasing effect size associated with increasing variability of the standards. However, this was likely because of the way MMN was calculated, subtracting the average response to the (variable) standards from the response to the deviants. Because increasing standard variability may result in some MMN occurring in the responses to standard tones, this serves to reduce the apparent MMN in the difference wave. This was also noted by Winkler and colleagues (1990), who found a significant effect of variability on the average standard response. Note that, in the present case, we did not compare the deviants to the standards but to comparable sounds in the control condition, alleviating this concern. Nevertheless, we took into account the possibility that the size of the chroma MMN effect might be smaller than that of the pitch height condition by designing the study with higher power than typical MMN studies. Our study included many (highly qualified) participants, we verified that all participants could make the relevant discrimination, and we replicated the null effect in two different studies. Although in previous studies using variable standards only ~10 participants were included, our study included 58 participants, divided into two similar experiments with more than 20 participants in each.

The possible effect of standard variability highlights the fact that chroma perception requires a higher level of abstraction than the perception of height—sounds with different heights may share the same chroma. This could explain why processing the dimension of chroma is not as automatic but rather likely requires higher cognitive processes.

Conclusion

Our results indicate that at the level of preattentive, automatic processing, pitch height is represented, whereas there is no evidence for similar representation of chroma, even in trained musicians. Processing chroma might require higher cognitive processes, such as attention, working memory, and learning. We suggest that octave equivalence of pure tones is not a low-level perceptual property but is rather a learned association. Our results do not support the notion of attention-independent neural representations specifically encoding chroma.

Acknowledgments

We are thankful to Prof. Roni Granot for fruitful discussions. We thank Assaf Brown for helping with recruitment of participants

from the music academy and for consulting musical issues regarding study design. We also thank Noam Segel for aiding with data collection and analysis of Experiment 2. We thank all research assistants who helped with data collection and analysis—Geffen Markusfeld, Michal Rabinovits, Eden Krispin, Anael Benistri, and Lior Matityahu, who helped with formatting bibliography. T. I. R. was supported by the Hoffman Leadership and Responsibility Program at the Hebrew University. I. N. was supported by a grant from the Israel Academy of Sciences (390/13). L. Y. D. is supported by Jack H. Skirball research fund.

Reprint requests should be sent to Tamar I. Regev, The Edmond and Lily Safra Center for Brain Science, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 91904, Israel, or via e-mail: tamaregev@gmail.com.

REFERENCES

- Aldwell, E., & Cadwallader, A. (2018). *Harmony and voice leading*. Boston, MA: Cengage Learning.
- Allen, D. (1967). Octave discriminability of musical and non-musical subjects. *Psychonomic Science*, *7*, 421–422.
- ANSI. (1994). *American National Standard Acoustical Terminology ANSI S1 1-1994*. New York: American National Standards Institute.
- Blackwell, H. R., & Schlosberg, H. (1943). Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology*, *33*, 407–419.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Briley, P. M., Breakey, C., & Krumbholz, K. (2013). Evidence for pitch chroma mapping in human auditory cortex. *Cerebral Cortex*, *23*, 2601–2610.
- Butler, J. S., Molholm, S., Fiebelkorn, I. C., Mercier, M. R., Schwartz, T. H., & Foxe, J. J. (2011). Common or redundant neural circuits for duration processing across audition and touch. *Journal of Neuroscience*, *31*, 3400–3406.
- Cynx, J. (1993). Auditory frequency generalization and a failure to find octave generalization in a songbird, the European starling (*Sturnus vulgaris*). *Journal of Comparative Psychology*, *107*, 140–146.
- Daikhin, L., & Ahissar, M. (2012). Responses to deviants are modulated by subthreshold variability of the standard. *Psychophysiology*, *49*, 31–42.
- Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America*, *76*, 57–66.
- Deouell, L. Y., Parnes, A., Pickard, N., & Knight, R. T. (2006). Spatial location is accurately tracked by human auditory sensory memory: Evidence from the mismatch negativity. *European Journal of Neuroscience*, *24*, 1488–1494.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781.
- Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H.-J., et al. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, 6755–6760.
- Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., & Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology*, *94*, 4269–4280.
- Gomes, H., Ritter, W., & Vaughan, H. G., Jr. (1995). The nature of preattentive storage in the auditory system. *Journal of Cognitive Neuroscience*, *7*, 81–94.
- Grimm, S., & Schröger, E. (2005). Pre-attentive and attentive processing of temporal and frequency characteristics within long sounds. *Cognitive Brain Research*, *25*, 711–721.
- Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H., & Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Animal Cognition*, *16*, 599–609.
- Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, & Psychophysics*, *74*, 1742–1760.
- Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*, 723–727.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*, 163–178.
- Kallman, H. J. (1982). Octave equivalence as measured by similarity ratings. *Perception & Psychophysics*, *32*, 37–49.
- Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*, 179–190.
- Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 579–594.
- Langner, G., & Ochse, M. (2006). The neural basis of pitch and harmony in the auditory system. *Musicae Scientiae*, *10*, 185–208.
- Loewy, D. H., Campbell, K. B., & Bastien, C. (1996). The mismatch negativity to frequency deviant stimuli during natural sleep. *Electroencephalography and Clinical Neurophysiology*, *98*, 493–501.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., et al. (2013). Processing of natural sounds: Characterization of multipeak spectral tuning in human auditory cortex. *Journal of Neuroscience*, *33*, 11888–11898.
- Moerel, M., De Martino, F., Santoro, R., Yacoub, E., & Formisano, E. (2015). Representation of pitch chroma by multi-peak spectral tuning in human auditory cortex. *Neuroimage*, *106*, 161–169.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590.
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, *115*, 140–144.
- Pakarinen, S., Huotilainen, M., & Näätänen, R. (2010). The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology*, *121*, 1043–1050.
- Poulin-Charronnat, B., Bigand, E., & Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: An event-related potential study. *Journal of Cognitive Neuroscience*, *18*, 1545–1554.
- Rosburg, T., Trautner, P., Dietl, T., Korzyukov, O. A., Boutros, N. N., Schaller, C., et al. (2005). Subdural recordings of the mismatch negativity (MMN) in patients with focal epilepsy. *Brain*, *128*, 819–828.
- Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials*, *62*, 437–448.

- Schnupp, J., Nelken, I., & King, A. (2011). *Auditory neuroscience: Making sense of sound*. Cambridge, MA: MIT Press.
- Schröger, E., & Wolff, C. (1996). Mismatch response of the human brain to changes in sound location. *NeuroReport*, 7, 3005–3008.
- Sergeant, D. (1983). The octave: Percept or concept. *Psychology of Music*, 11, 3–18.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, 36, 2346–2353.
- Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305–333.
- Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport*, 5, 844–846.
- Tervaniemi, M., Rytönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory*, 8, 295–300.
- Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, 372, 90–92.
- Wallin, N. L., Merker, B., & Brown, S. (Eds.) (2000). *The origins of music*. Cambridge, MA: MIT Press.
- Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences, U.S.A.*, 100, 10038–10042.
- Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1990). The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology*, 27, 228–235.
- Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, 129, 291–307.