



# Regularity extraction from non-adjacent sounds

Alexandra Bendixen<sup>1,2\*</sup>, Erich Schröger<sup>1</sup>, Walter Ritter<sup>3</sup> and István Winkler<sup>2,4</sup>

<sup>1</sup> Institute for Psychology, University of Leipzig, Leipzig, Germany

<sup>2</sup> Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary

<sup>3</sup> Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA

<sup>4</sup> Institute of Psychology, University of Szeged, Szeged, Hungary

## Edited by:

Micah M Murray, Université de Lausanne, Switzerland

## Reviewed by:

Elvira Brattico, University of Helsinki, Finland

Kai Alter, Newcastle University, UK

## \*Correspondence:

Alexandra Bendixen, Institut für Psychologie, Universität Leipzig, Seeburgstr. 14-20, D-04103 Leipzig, Germany

e-mail: alexan-

dra.bendixen@uni-leipzig.de

The regular behavior of sound sources helps us to make sense of the auditory environment. Regular patterns may, for instance, convey information on the identity of a sound source (such as the acoustic signature of a train moving on the rails). Yet typically, this signature overlaps in time with signals emitted from other sound sources. It is generally assumed that auditory regularity extraction cannot operate upon this mixture of signals because it only finds regularities between adjacent sounds. In this view, the auditory environment would be grouped into separate entities by means of readily available acoustic cues such as separation in frequency and location. Regularity extraction processes would then operate upon the resulting groups. Our new experimental evidence challenges this view. We presented two interleaved sound sequences which overlapped in frequency range and shared all acoustic parameters. The sequences only differed in their underlying regular patterns. We inserted deviants into one of the sequences to probe whether the regularity was extracted. In the first experiment, we found that these deviants elicited the mismatch negativity (MMN) component. Thus the auditory system was able to find the regularity between the non-adjacent sounds. Regularity extraction was not influenced by sequence cohesiveness as manipulated by the relative duration of tones and silent inter-tone-intervals. In the second experiment, we showed that a regularity connecting non-adjacent sounds was discovered only when the intervening sequence also contained a regular pattern, but not when the intervening sounds were randomly varying. This suggests that separate regular patterns are available to the auditory system as a cue for identifying signals coming from distinct sound sources. Thus auditory regularity extraction is not necessarily confined to a processing stage after initial sound grouping, but may precede grouping when other acoustic cues are unavailable.

**Keywords:** mismatch negativity (MMN), auditory processing, auditory object formation, sound grouping, integration, segregation, non-adjacent dependencies, implicit learning

## INTRODUCTION

Many auditory sources emit signals in a discontinuous manner over time, such as a walking person producing a series of footsteps. These discrete signals need to be bound together in order to form adequate representations of the sound sources in the environment (Bregman, 1990). Yet during the binding process, the auditory system has to avoid including intermittent signals that were emitted by other sources, such as another person's footsteps or someone coughing. This problem can be described as forming links between non-adjacent elements in a series of events. Current theories of auditory processing largely agree that the auditory system circumvents the formation of such non-adjacent links by initially grouping sounds on the basis of their individual features (e.g., pitch, timbre, perceived location), and then only forming links within the resulting groups (e.g., Sussman, 2005). The present study was designed to investigate whether the auditory system is capable of forming links between non-adjacent sounds under conditions that preclude grouping based on simple auditory features.

For studying the formation of links between non-adjacent sounds, a typical approach is to interleave two sound sequences in an alternating manner, and to have listeners perform a task on one of the sequences (e.g., Dowling et al., 1987). This task must be designed in such a way that it is only solvable when the listener succeeds in linking the relevant elements across the intervening elements from the other sequence. However, in such situations, listeners often use strategies to avoid the need for linking non-adjacent sounds (e.g., listeners can learn to specifically attend every other tone; see Dowling et al., 1987). Similarly, attention appears to emerge as the critical factor in implicit sequence learning studies testing the extraction of dependencies between non-adjacent stimuli (Pacton and Perruchet, 2008; Remillard, 2009). Thus it remains unclear whether non-adjacent stimuli can be linked without first forming groups (here mediated by attention) within which the originally non-adjacent stimuli become adjacent.

Event-related brain potentials (ERPs) offer a way to avoid attention-related grouping by using a "passive" version of the interleaved-sequences paradigm in which listeners do not perform

any task related to the sounds (Winkler, 1996; Sussman et al., 1999). To determine whether discrete sounds are linked together, one can set up some regularity between them, and insert occasional violations of this regularity into the sound sequence. If these regularity violations elicit the mismatch negativity (MMN) component of the ERP (Näätänen et al., 1978; for recent reviews, see Kujala et al., 2007; Näätänen et al., 2007), one can infer that the regularity violation was detected, and thus that the regularity must have been extracted (Schröger, 2007; Winkler, 2007). Finding that a regularity based on non-adjacent sounds has been extracted indicates that the relations between them have been explored. Thus in this paradigm, the presence of the MMN component is indicative of link formation between non-adjacent elements of a sound series (Winkler, 1996).

Applying this logic, a number of studies found no evidence that the auditory system would form links between non-adjacent tones in a series unless grouping was induced by a primary cue that distinguished the interleaved sequences (Sussman et al., 2001, 2007; Winkler et al., 2003a,b; Sussman and Steinschneider, 2006, 2009). In these studies, a repetition regularity (e.g., constant stimulus intensity) remained undetected when intervening tones with random intensities were inserted. Tone series were formed of the type, “CIICIIICIIICII...,” with “C” denoting the constant intensity value of the test sequence, “D” denoting a rare deviation from this constancy (i.e., the deviant stimulus used to probe MMN elicitation), and “I” denoting random intensity values of the other (intervening) sequence. No MMN was elicited when the two sequences differed only in their intensity values. MMN was regained when a second feature distinguishing the sequences (e.g., a difference in pitch) was introduced. These results were interpreted as showing a primacy of feature-based grouping over regularity extraction (Winkler, 1996; Sussman et al., 2001, 2007; Winkler et al., 2003a,b; Sussman and Steinschneider, 2006, 2009). Similar results were obtained when both the repetition regularity and the separation of the groups were based on pitch (Shinozaki et al., 2000), and when more complex regularities were employed (Sussman et al., 1998, 1999; Rahne et al., 2007; Rahne and Böckmann-Barthel, 2009). The auditory system thus does not appear to extract non-adjacent dependencies from a sound series unless sounds are first grouped on the basis of some distinguishing feature turning non-adjacent transitions into adjacent ones within each resulting group of sounds. This conclusion is again consistent with evidence from implicit sequence learning, in which the presence of other grouping cues assists in extracting non-adjacent dependencies (Creel et al., 2004). Other MMN studies also support the conclusion that regularity extraction does not operate on the individual sound input, but upon sound groups formed on the basis of simple auditory features (Ritter et al., 2000; Müller et al., 2005). Moreover, behavioral data obtained with structurally similar designs suggest that the auditory system does not detect non-adjacent transitions (Kaernbach and Demany, 1998).

Thus, a consistent picture emerges from the literature, suggesting that sounds are first grouped on the basis of feature similarity (Moore and Gockel, 2002), and that the resulting groups are then analyzed in terms of relations between the sounds, possibly detecting regularities between them (Sussman, 2005). This view gains additional plausibility by the fact that feature-based grouping

starts very early within the auditory pathway (Pressnitzer et al., 2008). The grouping feature is not necessarily pitch (Vliegen and Oxenham, 1999; Roberts et al., 2002; Neuhoff, 2003), but it must be a simple feature that can be assessed separately for each sound.

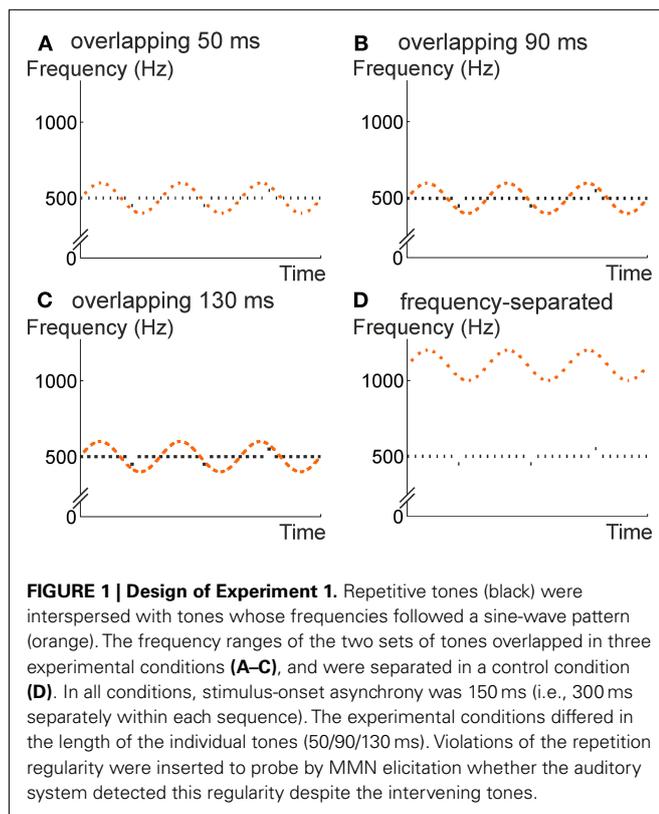
However, this view neglects the situation in which grouping on the basis of simple features is impossible because the interleaved sound sequences are not separated by such features. This situation has not been explored in detail in any of the previous studies. In fact, it has always been used as a control condition to be contrasted with another condition in which grouping would occur on the basis of separation in some readily available auditory feature. No attempt has been made to support non-adjacent link formation other than supporting initial grouping by introducing some feature difference between the interleaved sequences, by adding concurrent visual signals separating the two groups of sound, or by instructing participants to use some attentional strategy for group formation (see, e.g., Sussman et al., 1998; Rahne et al., 2007; Rahne and Böckmann-Barthel, 2009). We conducted two experiments aimed at revealing conditions that enable the auditory system to form links between non-adjacent tones without the need for grouping prior to link formation. Specifically, we investigated the influence of (1) sequence cohesiveness in the temporal domain, and (2) of regularity in the intervening sequence. As outlined above, non-adjacent link formation was assessed on the basis of MMN elicitation by deviations in one of the interleaved sequences comprising a simple repetition regularity (Winkler, 1996).

## EXPERIMENT 1

Experiment 1 was conducted as an exploratory test of two factors that we hypothesized to support link formation between non-adjacent tones. One of these factors, *sequence cohesiveness*, was manipulated within Experiment 1. The other factor, *regularity in the intervening sequence*, was introduced in Experiment 1 as a potentially critical difference from previous studies; its manipulation was left to Experiment 2.

By *sequence cohesiveness*, we refer to the effects of the inter-stimulus interval (ISI) separately within each of the two sequences. We hypothesized that for a given stimulus-onset asynchrony (SOA), links between non-adjacent tones would be easier to form when the ISI is shorter, as shorter gaps need to be bridged in this case. Shorter ISI values with constant SOA values were achieved by increasing tone duration.

Similarly to previous studies, we presented two interleaved sound sequences in an alternating manner. The *test* sequence was a repeating tone, except for occasional frequency deviations that were inserted to measure non-adjacent link formation (Winkler, 1996). Unlike previous studies, the *intervening* sequence was also regular. The frequency of the intervening tones oscillated around the common frequency of the tones in the repetitive test sequence according to a sine-wave pattern (**Figure 1** upper panels and lower left panel). We tested the extraction of the repetition regularity from the non-adjacent tones by inserting frequency deviants into the test sequence. Note that these deviants fell well within the frequency range covered by the intervening sequence. Thus they could not be identified as deviants without extracting the repetition regularity. The temporal cohesiveness of the sequence was



varied in three levels by manipulating the length of the tones within a constant rate of tone presentation (Figure 1).

For comparison with previous studies, we included a control condition in which the test and intervening sequences were separated in frequency (Figure 1 lower right panel). According to previous findings (Winkler, 1996; Sussman et al., 2001, 2007; Winkler et al., 2003a,b; Sussman and Steinschneider, 2006, 2009), frequency-based grouping should change the adjacency of the tones in this case, and it should be easy for the system to extract the repetition regularity from the test sequence.

All experimental conditions were presented first with *passive listening* and then with an active deviance detection task (*active listening*) in order to assess whether the information available to the MMN system would also be accessible to voluntary processing. In order to facilitate task performance, there was a short induction period of the test sequence alone in the *active listening* blocks, after which the intervening sequence faded in with a gradual intensity increase. Exemplary audio clips for each condition are provided as Supplementary Material.

## MATERIAL AND METHODS

### Participants

Fifteen healthy volunteers (eight male, two left-handed; mean age 22.1 years) participated in the experiment. All participants had frequency thresholds not higher than 20 dB hearing level (HL) in the 250–4000 Hz range and no threshold difference exceeding 10 dB between the two ears (assessed with a Mediroll, SA-5 audiometer). None of the participants were taking any medication affecting the central nervous system. Prior to the beginning

of the experiment, written informed consent was obtained from each participant according to the Declaration of Helsinki after experimental procedures and aims were explained to them. The study was approved by the Ethical Committee of the Institute for Psychology, Hungarian Academy of Sciences.

### Apparatus and stimuli

Participants were seated in an acoustically shielded chamber. A computer screen was placed in front of them at a distance of 1 m. Sinusoidal tones with an intensity of 50 dB sensation level (above hearing threshold, adjusted individually for each participant) were presented binaurally via headphones in a continuous series with a SOA of 150 ms. During the *active listening* blocks, participants held a response button in their dominant hand and responded with their thumb.

The tone series consisted of two interleaved sequences (a *test sequence* and an *intervening sequence*) presented in a strictly alternating schedule. The test sequence consisted of repeating pure tones (standards; 93%) with a frequency of 500 Hz and frequency deviants (7%) with a frequency of either 475 or 525 Hz, whichever was closer to the frequency of the preceding tone in the intervening sequence. Deviants were pseudorandomly distributed across the test sequence, with the restriction of a minimum inter-deviant interval of 2.7 s. The intervening sequence was composed of gliding tones whose frequency course followed the track of a slow sine-wave oscillation with a phase length of 3.4 s (see Figure 1). In three conditions (*overlapping-50*, *overlapping-90*, and *overlapping-130*), the mean frequency of the slow oscillation was identical to the repeating frequency of the test sequence (i.e., 500 Hz), and the oscillation covered a range of 450–550 Hz. In the fourth condition (*frequency-separated-50*), the mean frequency of the slow oscillation was 1100 Hz, and its range was 1000–1200 Hz. Within each condition, all tones of both sequences had the same duration. Duration values differed between conditions, taking the values of 50, 90, or 130 ms as indicated by condition names. All tones included raised-cosine ramps with 5 ms rise and 5 ms fall times. Conditions were presented in blocks of 2110 stimuli (1055 per sequence, including 75 deviants). Stimulation was randomized individually for each participant.

In the *passive listening* version, both sequences were presented at full and equal amplitude throughout the block. In the *active listening* version, the test sequence started and remained at full amplitude, while the intervening sequence was absent for the initial 20 tones and then gradually reached full amplitude from tone 21 to 100 in linear steps of 1/80. The rationale of this procedure was to present participants with an induction period to facilitate hearing out the test sequence.

### Procedure

For each condition, three blocks were administered while participants watched a silent, subtitled movie, and were instructed to ignore the tones (*passive listening*; 225 deviant stimuli per condition). After all *passive listening* blocks were completed, each condition was ran once more in a separate active block in which participants were instructed to respond to deviants in the test sequence by button presses as fast and accurately as possible (*active listening*; 75 deviant stimuli per condition).

Training blocks were given between the passive and active parts to familiarize participants with the task. The order of conditions was randomized for each participant, separately for the *passive* and *active listening* parts. Each block lasted 5.3 min. The overall duration of the experiment amounted to 85 min net time. Breaks between blocks were given to participants as needed.

### Behavioral data recording and analysis

In the *active listening* part, button presses made by participants were continuously recorded. Responses during the initial 30 s were discarded from the analysis because performance was facilitated by the reduced intensity of the intervening sequence (see above). The remaining 4.8 min comprised 956 tones of the test sequence with 68 frequency deviants amongst them. Any response given within an interval of 150–1000 ms after a deviant was counted as a *hit*. The proportion of hits was calculated as

$$p_H = \frac{N_{\text{hits}}}{N_{\text{deviants}}} \quad (1)$$

with  $N$  being the number of times that the event was observed or presented. Any response given outside the specified response windows was counted as a *false alarm*. Calculating the proportion of false alarms in paradigms with fast-paced stimulus presentation requires modification of the conventional formula because responses cannot unambiguously be related to a particular preceding stimulus. Based on the procedure suggested by Bendixen and Andersen (submitted), we calculated the false alarm rate as

$$p_{FA} = \frac{N_{\text{false alarms}} \times T_{\text{Response}}}{T_{\text{Sequence}} - (N_{\text{deviants}} \times T_{\text{Response}})} \quad (2)$$

where  $T_{\text{Response}}$  denotes the length of the response window, and  $T_{\text{Sequence}}$  the duration of the entire sequence. This formula ensures that *hits* and *false alarms* are probed in comparable time intervals, which makes the method independent from the specific choice of the response window (Bendixen and Andersen, submitted).

The sensitivity index  $d'$  was then derived from the hit and false alarm rates according to signal detection theory (SDT; Green and Swets, 1966; Macmillan and Creelman, 2005). Calculations were based on the Gaussian model assuming homogeneous variances, with

$$d' = Z(p_H) - Z(p_{FA}) \quad (3)$$

where  $p_H$  is the hit rate,  $p_{FA}$  is the false alarm rate, and  $Z$  is the inverse of the cumulative normal distribution. To avoid infinite values in the cumulative normal distribution, hit and false alarm rates were adjusted to  $1-1/(2N)$  when they were actually 1 and to  $1/(2N)$  when they were actually 0, with  $N$  being the number of observation periods (Snodgrass and Corwin, 1988; Stanislaw and Todorov, 1999; Macmillan and Creelman, 2005).

The  $d'$  values were compared between the four conditions in a repeated-measures analysis of variance (ANOVA) with the factor condition (four levels: *overlapping-50*, *overlapping-90*, *overlapping-130*, *frequency-separated-50*).

### Electrophysiological data recording and analysis

Electrophysiological recordings were made continuously with Ag/AgCl electrodes during the *passive* and *active listening* parts. Electrodes were placed at F3, Fz, F4, C3, Cz, C4, and Pz according to the international 10–20 system (Jasper, 1958). Additional electrodes were placed at the tip of the nose, which served as a reference, and at the left and right mastoid sites. Eye movements were monitored by electrodes placed above and below the left eye and at the outer canthi of both eyes, which were bipolarized off-line to yield vertical and horizontal electroocular activity (EOG). EEG and EOG signals were amplified (0–40 Hz) by SynAmps amplifiers (Neuroscan Inc.), sampled at 250 Hz, and filtered off-line using a 0.5–30 Hz band pass filter.

For each trial, an epoch of 500 ms duration including a 50 ms pre-stimulus baseline was extracted from the continuous EEG record to form ERPs. Epochs with amplitude changes exceeding  $100 \mu\text{V}$  on any channel were rejected from further analysis, which led to retaining 86.5% of the epochs on average. Epochs for standards and deviants from the test sequence were averaged separately for the *passive* and *active listening* versions of each condition. One standard stimulus before and three standard stimuli after each deviant were excluded from averaging to avoid contaminating the standard ERP averages by indicators of processing the deviant, and because standards immediately after a deviant sound may elicit a small mismatch response relative to standards preceded by other standards (Sams et al., 1984; Nousak et al., 1996).

Separately for each condition of the *passive* and *active listening* parts, deviant-minus-standard difference waves were formed in order to identify the MMN component as an indicator of detecting the frequency deviants. Based on the first notable negative component in the difference waveforms, mean ERP amplitudes were measured in the interval from 120 to 160 ms in all conditions, and tested against zero using one-sample, one-tailed Student's  $t$ -tests to verify the presence of the MMN component. In order to increase signal-to-noise ratio, data at Fz were re-referenced against the average signal at the two mastoids (Schröger, 2005). All ERP analyses were performed on the re-referenced data at Fz. Separately for the *passive* and *active listening* parts, mean MMN amplitudes were compared across the four conditions in repeated-measures ANOVAs with the factor Condition (four levels: *overlapping-50*, *overlapping-90*, *overlapping-130*, *frequency-separated-50*).

All significant ANOVA effects are reported with the partial  $\eta^2$  effect size measure. The Greenhouse–Geisser correction (Greenhouse and Geisser, 1959) was applied when the assumption of sphericity was violated. *Post hoc* tests for statistical analyses were carried out with the Bonferroni correction of the confidence level for multiple comparisons.

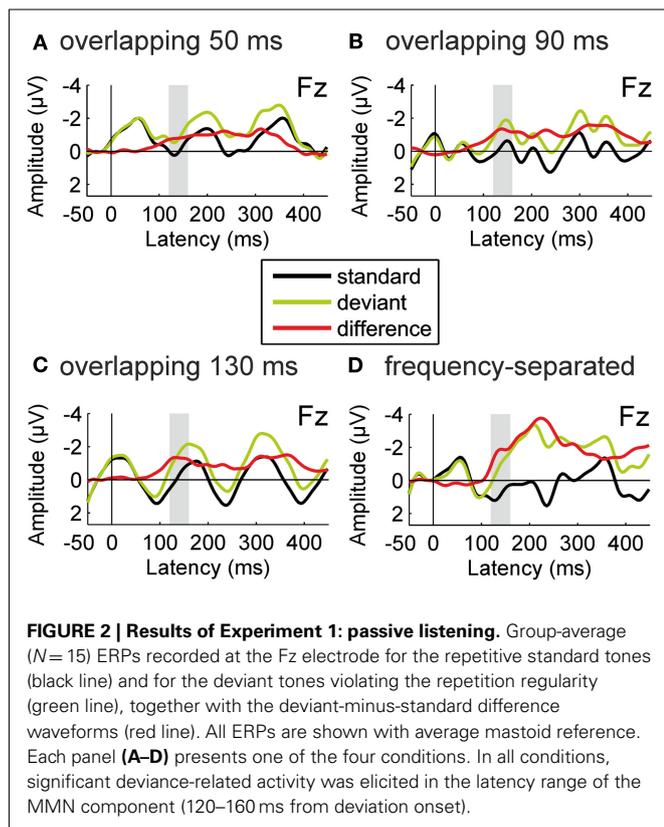
### RESULTS AND DISCUSSION

In the *active listening* part, deviants were detected with a mean sensitivity of 0.279 (*overlapping-50*), 0.186 (*overlapping-90*), 0.419 (*overlapping-130*), and 4.122 (*frequency-separated-50*). The ANOVA showed a significant effect of condition,  $F(3, 42) = 144.822$ ,  $p < 0.001$ ,  $\epsilon_{GG} = 0.450$ ,  $\eta^2 = 0.912$ . *Post hoc* tests revealed significant differences between the *frequency-separated-50* condition and each of the other conditions (all  $p$  values  $< 0.001$ ).

None of the pair-wise comparisons between the overlapping conditions with different tone durations were significant (all  $p$  values  $>0.569$ ).

**Figure 2** shows the ERPs obtained for the standard and deviant tones in the test sequence during *passive listening* as well as the deviant-minus-standard difference waveforms, separately for the four conditions. **Figure 3** displays the difference waves of each condition for the whole electrode set with the original nose reference. In all conditions, a significant negative deflection was elicited in the latency range of the MMN component with maximum at Fz and polarity inversion at the mastoid electrodes (all  $p$  values  $<0.01$  for Fz re-referenced to average mastoids). Thus deviations from the repetition regularity in the test sequence were detected by the auditory system irrespective of whether the intervening sequence was separated from or overlapping the frequency range of the test sequence. The amplitude of the MMN component during *passive listening* was not modulated by condition,  $F(3, 42) = 1.606$ ,  $p = 0.202$ .

**Figure 4** displays the ERPs obtained for the standard and deviant tones in the test sequence during *active listening* as well as the deviant-minus-standard difference waveforms, separately for the four conditions. Again, significant negative deflections were elicited in the MMN latency range in all conditions (all  $p$  values  $<0.05$ ). Amplitudes in the MMN latency range were influenced by condition,  $F(3, 42) = 6.347$ ,  $p < 0.01$ ,  $\eta^2 = 0.312$ . *Post hoc* tests revealed significant differences between the *frequency-separated-50* condition and each of the other conditions (all  $p$  values  $<0.05$ ), whereas none of the other conditions differed from each other



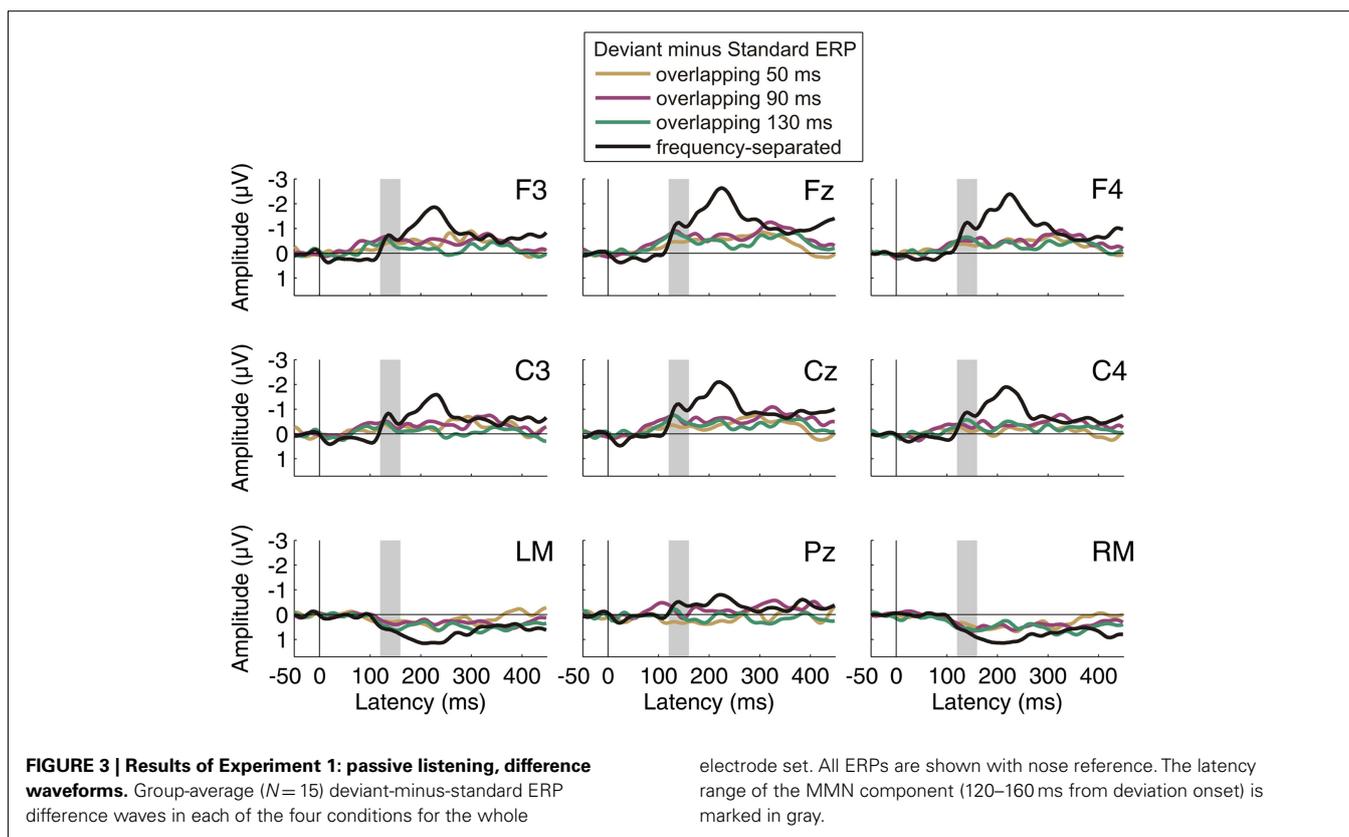
(all  $p$  values  $>0.999$ ). The morphology of the ERPs following the MMN latency range (**Figure 4**), together with the behavioral data, suggest that the condition effect on the activity in the MMN latency range during *active listening* was mainly caused by an overlapping N2 component whose amplitude varies with conscious deviance detection (Novak et al., 1990).

The amplitude of the MMN component in the *passive listening* part, and probably also in the *active listening* part, was not modulated by condition (although there is a marked difference between conditions in a later latency range). This result has two important implications. First, in contrast to our expectations, the temporal cohesiveness of the sequence had no impact on link formation between the tones in the sequence. This finding will be addressed in the General Discussion. Second, and more importantly, the presence or absence of a frequency difference between two interleaved sound sequences had no impact on the elicitation nor on the amplitude of the MMN component to deviations in one of the sequences. This is in contrast with a number of previous findings (Winkler, 1996; Shinozaki et al., 2000; Sussman et al., 2001, 2007; Winkler et al., 2003a,b; Sussman and Steinschneider, 2006, 2009). Following the logic outlined in the Introduction, the results of Experiment 1 could be interpreted as indicating the formation of links between the non-adjacent tones in the test sequence. There are, however, some alternative interpretations that cannot be ruled out on the basis of Experiment 1 alone.

One alternative interpretation posits that the two sound sequences were distinguished on the basis of their static vs. dynamic frequency characteristics. The tones of the test sequence were of constant frequency, while the tones of the intervening sequence were gliding in frequency. We cannot exclude that this difference between the two streams can be used as a primary acoustic cue for auditory stream segregation (although we are not aware of any evidence supporting this possibility).

The second alternative interpretation is based on the fact that following the sine-wave trajectory of the intervening tones, there were regular segments of 3–4 tones in which the frequency of the intervening tones was more than one semitone above or below that of the test tones. Although unlikely, a frequency difference of one semitone at 150 ms SOA might induce stream segregation (van Noorden, 1975). Thus for short portions of the tone series, the two sequences might have been segregated on the basis of their local frequency difference. Given that repetition regularities are extracted after just one or two repetitions (Horváth et al., 2001, 2008b; Haenschel et al., 2005; Bendixen et al., 2007), it is possible that deviants occurring at the end of such short segregated segments elicited the MMN component, and that these local deviations gave rise to the observed significant MMN response.

A third alternative is that the auditory system does not capture the repetition regularity at all, but responds to the probabilities of the various frequencies in the tone series, based on differential refractoriness of frequency-specific neurons (Walker et al., 2001; May and Tiitinen, 2010). According to this argument, any tone whose frequency sufficiently differs from that of the standard (500 Hz) should elicit a negativity relative to the frequently occurring standard tones. The MMN-like component measured in Experiment 1 would then rather reflect a difference in the amplitude of the N1 component elicited by tones appearing with



different probabilities. Such “N1 contamination” of the MMN response has been shown for frequency deviations as small as 2% (Horváth et al., 2008a), which is well below the 5% deviation between deviants and standards chosen here.

In conclusion, Experiment 1 may suggest that the auditory system is able to extract a repetition regularity from non-adjacent tones, but a number of alternative interpretations cannot be ruled out. Experiment 2 was designed to clarify this issue.

## EXPERIMENT 2

Experiment 2 was conducted to replicate the findings of Experiment 1 while contrasting the alternative interpretations. Our strategy was to set up a condition in which connecting non-adjacent sounds would become more difficult, whereas all previously described alternative interpretations would still predict a difference between the deviant- and standard-stimulus responses. Finding no or a significantly reduced deviant-minus-standard difference under these conditions would indicate that the deviant-minus-standard difference found in Experiment 1 can be interpreted as demonstrating link formation between non-adjacent tones.

The factor that we manipulated in Experiment 2 was the amount of *regularity in the intervening sequence*. The intervening sequence being regular rather than random was introduced in Experiment 1 as a potentially critical difference from previous studies. In order to explicitly test the impact of this modification, Experiment 2 comprised one of the conditions from Experiment 1 (*overlapping-130*) and a control condition which was identical to

it except for the order of the intervening sounds being randomized (Figure 5). Note that all alternative explanations for the elicitation of the deviant-minus-standard difference outlined above have been preserved by this modification. Exemplary audio clips for each condition are provided as Supplementary Material.

## MATERIAL AND METHODS

### Participants

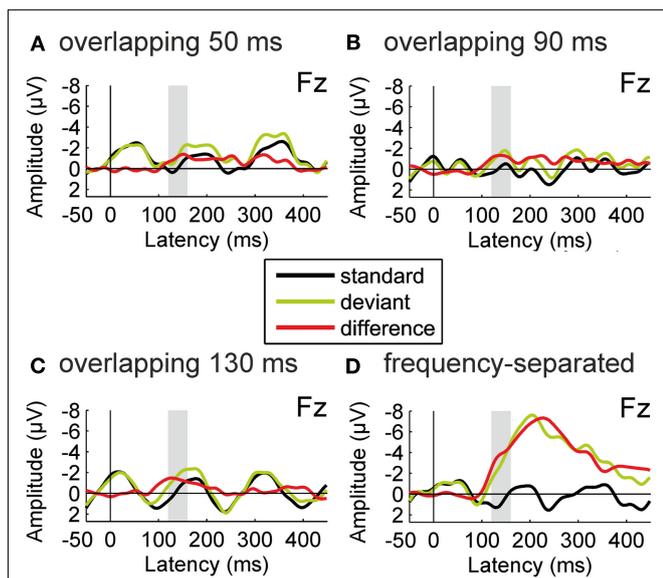
Sixteen healthy volunteers (six male, two left-handed; mean age 20.9 years) participated in the experiment. The same inclusion criteria and ethics procedure were applied as in Experiment 1.

### Apparatus and stimuli

Apparatus and stimuli were identical to those in Experiment 1 except for the following modifications. Only the *overlapping-130* condition (with 130 ms duration for all tones of both sequences) was administered; it is termed *intervening-regular* condition in this experiment. A second, *intervening-random* condition was derived from the *intervening-regular* condition by randomizing the order of the gliding tones so that the frequency envelope of the intervening tone sequence no longer followed a sine-wave pattern (Figure 5).

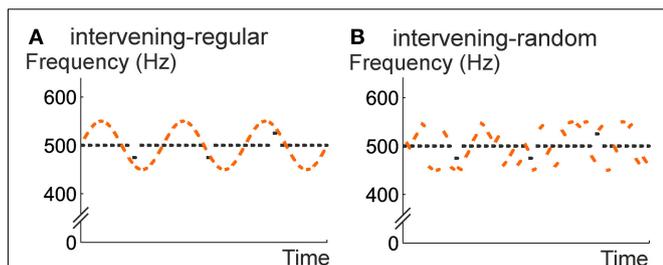
### Procedure

Five blocks for each of the two conditions were administered while participants watched a silent, subtitled movie and were instructed to ignore the tones (*passive listening*; 375 deviant stimuli per condition). The order of the *passive listening* blocks was randomized



**FIGURE 4 | Results of Experiment 1: active deviance detection.**

Group-average ( $N = 15$ ) ERPs recorded at the Fz electrode for the repetitive standard tones (black line) and the deviant tones violating the repetition regularity (green line), together with the deviant-minus-standard difference waveforms (red line). All ERPs are shown with average mastoid reference. Each panel (A–D) presents one of the four conditions. In all conditions, significant deviance-related activity was elicited in the latency range of the MMN component (120–160 ms from deviation onset). Note that the amplitude scale differs from that of Figures 2 and 3.



**FIGURE 5 | Design of Experiment 2.** The “overlapping 130 ms” condition of Experiment 1 was repeated (A) and complemented by a control condition in which the tones of the sine-wave envelope appeared in random order (B). Again, violations of the repetition regularity were inserted to probe whether the auditory system detected the regularity across the intervening sounds.

separately for each participant with the restriction of no more than two blocks of the same condition in a row. After all *passive listening* blocks were completed, two *active listening* blocks (one for each condition, containing 75 deviant stimuli) were administered in which participants were instructed to respond to deviants in the test sequence by button presses as fast and accurately as possible. Training blocks were given between the *passive* and *active listening* parts to familiarize participants with the task. Condition order of the two *active listening* blocks was counterbalanced across participants. Finally, three more active blocks of the *intervening-regular* condition were administered whose results were used for

validating the new procedure of analyzing behavioral data in a fast-paced signal detection paradigm (Bendixen and Andersen, submitted). Each block lasted 5.3 min. The overall duration of the experiment amounted to 79.5 min net time. Breaks between blocks were given to participants as needed.

#### Behavioral data recording and analysis

Participants' button presses were continuously recorded, and deviant detection performance was analyzed as described for Experiment 1 based on the initial two *active listening* blocks (one for each condition). The  $d'$  values in the two conditions (*intervening-regular*, *intervening-random*) were compared using a paired, two-tailed Student's  $t$ -test.

#### Electrophysiological data recording and analysis

Electrophysiological data recording and analysis were identical to those in Experiment 1 except for the following modifications. After artifact removal, 91.2% of the epochs were retained on average. For a control analysis with respect to the alternative interpretation based on differential refractoriness of frequency-specific neurons, not only epochs for standards and deviants from the test sequence were averaged but also epochs for tones from the intervening sequence whose mean frequency was between 472.5 and 477.5 Hz (around the 475 Hz deviant tone) or between 522.5 and 527.5 Hz (around the 525 Hz deviant tone).

Only ERP data from the *passive listening* part are reported here. Based on the first notable negative component, mean ERP amplitudes of the deviant-minus-standard difference waveforms were measured in the interval of 135 to 175 ms in both conditions, and tested against zero using one-sample, one-tailed Student's  $t$ -tests to verify the presence of the MMN component. In order to increase signal-to-noise ratio, data at Fz were re-referenced against the average signal at the two mastoids (Schröger, 2005). All ERP analyses were performed on the re-referenced data at Fz. Mean MMN amplitudes were compared between the two conditions (*intervening-regular*, *intervening-random*) in a paired, two-tailed Student's  $t$ -test. Mean ERP amplitudes of the selected intervening tones were compared against standard ERPs and deviant ERPs in the latency range of the MMN component by means of paired, two-tailed Student's  $t$ -tests.

## RESULTS AND DISCUSSION

During *active listening*, deviants were detected with a mean sensitivity of 0.275 (*intervening-regular*) and 0.204 (*intervening-random*). There was no significant difference in detection performance between the two conditions,  $t(15) = 0.719$ ,  $p = 0.483$ .

Figure 6 shows the ERPs obtained for the standard and deviant tones during *passive listening* as well as the deviant-minus-standard difference waveforms, separately for the two conditions. In both conditions, a significant negative deflection was elicited in the latency range of the MMN component [*intervening-regular*,  $t(15) = 4.590$ ,  $p < 0.001$ , *intervening-random*,  $t(15) = 2.496$ ,  $p < 0.05$ ]. Thus deviations from the repetition regularity in the test sequence were differentially processed by the auditory system despite the presence of the intervening tones.

The amplitude of the deviant-minus-standard difference was significantly modulated by condition,  $t(15) = 6.230$ ,  $p < 0.05$ ,

$\eta^2 = 0.293$ , with higher amplitudes obtained in the *intervening-regular* than in the *intervening-random* condition. This result cannot be explained on the basis of any of the alternative interpretations given above. First, the difference between the two sequences in terms of static (constant) vs. dynamic (gliding) frequency characteristics of the tones is present in both conditions. Second, both conditions contain short segments of the tone series in which the two sequences could be segregated on the basis of their local frequency difference. Third, the global probabilities for the different frequencies within the tone series are identical in the two conditions.

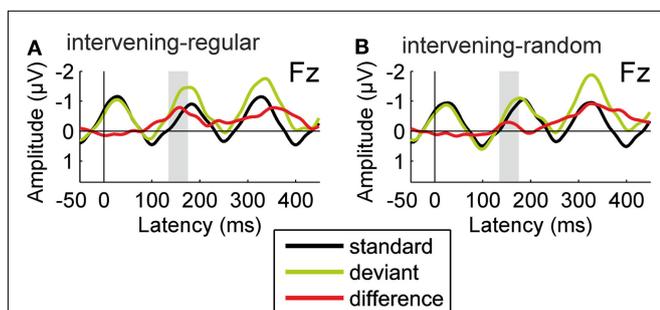
One may, however, argue that there are still differences in the local probabilities of the different tones, and that the auditory system may be sensitive to these local characteristics. This is a local version of the interpretation based on differential refractoriness of frequency-specific neurons (Walker et al., 2001; May and Tiitinen, 2010). To test this explanation, we compared the ERPs elicited by the standard and deviant tones of the test sequence with those elicited by the intervening tone glides whose mean frequency was equal or very similar to the frequency of the deviant tones ( $\pm 0.5\%$ ). The results of this analysis are shown in **Figure 7**. For the *intervening-regular* condition, the ERPs elicited by intervening tones sharing the deviant frequency do not differ from the ERPs elicited by standard tones,  $t(15) = -0.576$ ,  $p = 0.573$ , whereas they significantly differ from the deviant ERPs in the MMN latency range,  $t(15) = 4.132$ ,  $p < 0.001$ . For the *intervening-random* condition, the ERPs elicited by the intervening tones differ from the standard-tone ERPs in an early latency range. However, in the MMN latency range, the ERPs elicited by the intervening tones differ neither from those elicited by the standard tones,  $t(15) = 0.758$ ,  $p = 0.460$ , nor from those elicited by the deviant tones,  $t(15) = 0.700$ ,  $p = 0.495$ . These results suggest that the small deviant-minus-standard difference found in the MMN latency range in the *intervening-random* condition probably mainly reflects differential

refractoriness of frequency-specific neurons. In contrast, the deviant-minus-standard difference obtained in the *intervening-regular* condition cannot be explained on the basis of refractoriness as it is only present for tones at the deviant frequency occurring unexpectedly, but not for tones at the deviant frequency occurring as part of the sine-wave trajectory. It thus seems likely that the MMN in the *intervening-regular* condition reflects deviance detection based on the extraction of the regularity in the test sequence.

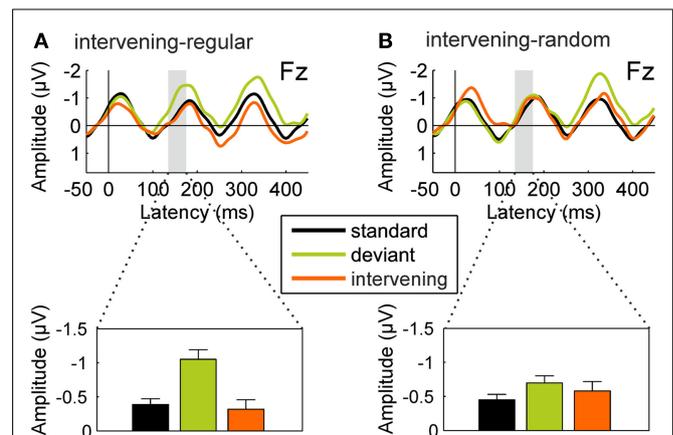
Taken together, the results of Experiment 2 replicate those of Experiment 1 while excluding a range of alternative explanations. Thus MMN elicitation by deviants in the test sequence can be taken to indicate that the underlying regularity was extracted despite the presence of the intervening tones. The extraction of a regularity carried by non-adjacent tones received strong support from the intervening sequence having its own regularity (which was different from that of the test sequence).

## GENERAL DISCUSSION

The present experiments were aimed at finding conditions under which the auditory system would form links between non-adjacent elements in a tone series without prior grouping by simple auditory features. Our results support the hypothesis that the auditory system indeed possesses the capability of linking non-adjacent tones. We found that one supportive condition for this capability is to intersperse two tone sequences that both carry regular characteristics, rather than to intersperse one regular and one random tone sequence.



**FIGURE 6 | Results of Experiment 2.** Group-average ( $N = 16$ ) ERPs recorded at the Fz electrode for the repetitive standard tones (black line) and the deviant tones violating the repetition regularity (green line), together with the deviant-minus-standard difference waveforms (red line). All ERPs are shown with average mastoid reference. In the *intervening-regular* condition (**A**), significant deviance-related activity was elicited in the latency range of the MMN component (135–175 ms from deviation onset). A significant negative difference in the same latency range was also observed in the *intervening-random* condition (**B**), but with significantly smaller amplitude.



**FIGURE 7 | Refractoriness control analysis for Experiment 2.** Upper panels: Group-average ( $N = 16$ ) ERPs recorded at the Fz electrode for the repetitive standard tones (black line), deviant tones violating the repetition regularity (green line), and intervening tone glides matching the deviant frequency (orange line). All ERPs are shown with average mastoid reference. Lower panels: Mean ERP amplitudes in the latency range of the MMN component (135–175 ms from deviation onset). In the *intervening-regular* condition (**A**, left column), the intervening tone glides do not seem to be processed differently from the standard tones, arguing against a contribution of refractoriness to the negativity elicited by the deviants in the MMN latency range. In the *intervening-random* condition (**B**, right column), there are early differences between standard and intervening tones which, however, do not translate into the MMN latency range.

Results of the two experiments demonstrate that occasional frequency deviations embedded in a sequence of repetitive frequencies are detected by the auditory system (as evidenced by the elicitation of the MMN component) even in the presence of intervening tones covering the same frequency range. Following the measurement logic of the MMN (Schröger, 2007; Winkler, 2007), this result suggests that the regularity of tone frequency repetition was extracted from non-adjacent sounds. Experiment 2 further qualified this finding by showing that non-adjacent link formation was facilitated when both interleaved sound sequences were following independent regularities, relative to when only one of them was regular while the other consisted of tones with random frequencies. This explains why many previous studies have found no evidence for regularity extraction from non-adjacent sounds. In these studies, the intervening tones were either randomly chosen (Winkler, 1996; Sussman et al., 2001, 2007; Winkler et al., 2003a,b; Sussman and Steinschneider, 2006, 2009; Rahne et al., 2007; Rahne and Böckmann-Barthel, 2009), or else they were regular but their pattern was not independent from that of the test sequence (Sussman et al., 1998, 1999; Shinozaki et al., 2000; Rahne and Sussman, 2009). Another contributing factor for the divergence of our finding from that of some previous studies (e.g., Shinozaki et al., 2000) may be that we used relatively short temporal intervals between the tones (150 ms for the entire tone series, 300 ms when considering each sequence separately), which has been shown to facilitate the formation of links between tones (Yabe et al., 1997, 1998; Winkler et al., 1998).

Thus our conclusion differs from many previous studies in that, in order to form links between non-adjacent sounds, it is not necessary to change the adjacency of the elements of the sound series by introducing a feature difference or by providing visual/attentional cues. The auditory system can detect that non-adjacent elements belong together based only on their regular behavior over time. It seems likely that this capability comes into play only when there is no possibility of grouping the initial auditory input on the basis of more readily available principles such as similarity in the primary features (Moore and Gockel, 2002). As soon as there are differences in the primary features, grouping based on these features probably precedes the operation of the regularity extraction mechanism, and determines the input upon which it operates (Ritter et al., 2000; Müller et al., 2005; Sussman, 2005).

Two unexpected aspects of the present dataset require some further discussion. First, temporal cohesiveness (manipulated through the length of ISI in Experiment 1) did not appear to influence the formation of links between non-adjacent tones. The MMN amplitude elicited during *passive* and *active listening* in Experiment 1 was not modulated by ISI. It is possible that link formation should not be regarded as “filling the gap” between the terminal part of one sound and the initial part of the next one, but as linking the initial sound parts. Compatible evidence comes from studies whose results suggest that the initial part of a sound plays a more important role in forming a sound representation than later sound parts (Grimm et al., 2004; Weise et al., 2007, 2010; Timm et al., 2011). Processes more reliant on ISI may come into play when the two sounds that must be linked are not identical (Bregman et al., 2000; but see Bee and Klump, 2005).

Second, ERP differences were found between standards and deviants beyond the MMN latency range in all conditions of both experiments. These differences were not expected, but can be *post hoc* interpreted as indications of further processing of the deviance. One possibility is that due to the complex stimulus configuration, a more intense update of the underlying model of the auditory regularities than usual occurred (Winkler and Czigler, 1998). Similar negativities following the MMN have been obtained in other complex auditory settings (Zachau et al., 2005; Horváth et al., 2009) and are sometimes referred to as late discriminative negativity (LDN). Alternatively, these late negativities could be purely refractoriness-related in the present study, as they were also observed in the *intervening-random* condition of Experiment 2 in which no MMN was elicited. Further studies should be conducted to clarify the conditions of elicitation of these late negativities, and to investigate the factors underlying their morphology (more sustained in Experiment 1, more transient in Experiment 2 of the present study). Importantly, regardless of their interpretation, these late negativities do not affect activity in the MMN latency range, and thus do not challenge our conclusions on link formation between non-adjacent sounds in the present tone series.

In theory, the capacity of the auditory system revealed here provides a powerful basis for disentangling a sound mixture based on underlying regular patterns. The non-adjacent links discovered by regularity extraction could immediately be used for auditory stream segregation and thus in turn for forming veridical representations of the sound sources in the environment (Bregman, 1990). The restriction of this capacity to situations in which two regular (rather than one regular and one random) sequences are mixed together may illustrate an ecologically plausible adaptation of the auditory system. The presence of regularities within both (all) sound sequences in a mixture is more consistent with real-life situations, in which two (or more) concurrently active sound sources are typically regular rather than artificially random. A supportive role of such regularities for decomposing the auditory input is in line with recent theoretical (Denham and Winkler, 2006; Winkler et al., 2009, 2012) and experimental work (Bendixen et al., 2010; Andreou et al., 2011; Rimmele et al., in press; Bendixen et al., in press) on the role of regular patterns in auditory scene analysis. It is also consistent with earlier suggestions from Jones and colleagues (Jones, 1976; Jones et al., 1981; Jones and Boltz, 1989), although our account is not based on attentional entrainment to the regularity.

In practice, however, it is not possible to conclude from the present set of results whether regularity extraction from (i.e., link formation between) non-adjacent sounds actually led to the segregation of these sounds into distinct perceptual streams. Although the information about the common regular properties of the tones was obviously available to the auditory system, it is unclear whether the MMN in this case reflects the actual perceptual organization, or just one of the perceptual alternatives (Horváth et al., 2001) that may not result in being the dominant percept (Winkler et al., 2012). Participants were not asked to report whether they perceived the tone series as integrated into one or splitting into two streams. The low performance in

the objective deviance detection task rather indicates that they had difficulties in following the test sequence as a separate stream unless there was an additional frequency difference between the putative streams (frequency-separated condition of Experiment 1). However, participants may have adopted a specific attentional strategy for trying to solve the task, such as listening out for every other tone in the series (Dowling et al., 1987), rather than relying on the outcome of their pre-attentive regularity extraction and deviance detection mechanism (which might have led to better performance). It is known from a number of other MMN studies with complex stimulus configurations that passive deviance detection (indicated by MMN elicitation) and active deviance detection (indicated by task performance) can dissociate, such that MMN is elicited in situations in which participants demonstrate poor behavioral performance even with full knowledge of the regularity (Paavilainen et al., 2001, 2007; Takegata et al., 2005; van Zuijen et al., 2006; Bendixen et al., 2008). Therefore the issue of whether perceptual stream segregation is possible on the basis of the present arrangement of regularities remains to be investigated.

Regardless of whether the information available to the MMN system is eventually used in auditory stream segregation, the present evidence supporting the notion of non-adjacent link formation is highly informative for MMN-related processing. The auditory system has been suggested to possess advanced capabilities for determining relations between consecutive sounds (Näätänen et al., 2001, 2010; Winkler, 2003; Conway et al., 2009; Furl et al., 2011), given that its input is essentially sequential in nature. This capability has been evidenced by MMN responses elicited by violations of complex inter-tone relations (Saarinen et al., 1992; Paavilainen et al., 2001, 2007; van Zuijen et al., 2006; Bendixen et al., 2008). By means of MMN, it has also been demonstrated that the auditory system is able to extract regularities from continuous sequences in which the tone relations keep dynamically changing (Sussman and Winkler, 2001; Haenschel et al., 2005; Bendixen et al., 2007, 2008; Bendixen and Schröger, 2008; Rahne and Sussman, 2009; Costa-Faidella et al., 2011). However, so far it has been assumed that all these complex and adaptive MMN responses are limited to dealing with relations between adjacent tones in a sequence. Relieving this computational constraint by showing that the MMN is responsive also to violations of non-adjacent transitions significantly increases the possibilities of modeling the auditory environment.

In conclusion, we demonstrate that the auditory system has the capacity to form links between non-adjacent elements in a tone series without relying on prior grouping by simple auditory features. Linking non-adjacent sounds was shown to occur when sounds with independent regular patterns were interspersed with each other. The capacity for non-adjacent link formation may be an important factor in finding the sound sources in typical auditory environments in which multiple sources are active at the same time.

## ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, BE 4284/1–1 to Alexandra Bendixen, SCH 375/20–1 to Erich Schröger) and by

the European Commission's Seventh Framework Programme (ICT-FP7-231168 to István Winkler). The experiments were realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN. The EEG data were analyzed with EEGLab (Delorme and Makeig, 2004). The authors thank Zsuzsanna D'Albini and Orsolya Száldy for collecting the data as well as Andreas Widmann for helpful discussion.

## SUPPLEMENTARY MATERIAL

The Audio Files for this article can be found online at [http://www.frontiersin.org/Auditory\\_Cognitive\\_Neuroscience/10.3389/fpsyg.2012.00143/abstract](http://www.frontiersin.org/Auditory_Cognitive_Neuroscience/10.3389/fpsyg.2012.00143/abstract)

### Audio File S1 | Experiment 1, passive listening, overlapping-50 condition.

Excerpt of the tone series presented in Experiment 1 during passive listening with overlapping frequencies of the test and intervening sequences and 50 ms duration of each individual tone.

### Audio File S2 | Experiment 1, passive listening, overlapping-90 condition.

As Audio File S1, but with 90 ms duration of each individual tone.

### Audio File S3 | Experiment 1, passive listening, overlapping-130 condition.

As Audio File S1, but with 130 ms duration of each individual tone.

### Audio File S4 | Experiment 1, passive listening, frequency-separated-50 condition.

As Audio File S1, but with frequency separation between the test and intervening sequences.

### Audio File S5 | Experiment 1, active listening, overlapping-50 condition.

Excerpt of the tone series presented in Experiment 1 during active listening with overlapping frequencies of the two tone sequences and 50 ms duration of each individual tone. In contrast to the passive listening version, the intervening sequence is initially absent and then fades in with a gradual intensity increase to facilitate hearing out the test sequence.

### Audio File S6 | Experiment 1, active listening, overlapping-90 condition.

As Audio File S5, but with 90 ms duration of each individual tone.

### Audio File S7 | Experiment 1, active listening, overlapping-130 condition.

As Audio File S5, but with 130 ms duration of each individual tone.

### Audio File S8 | Experiment 1, active listening, frequency-separated-50 condition.

As Audio File S5, but with frequency separation between the test and intervening sequences.

### Audio File S9 | Experiment 2, passive listening, intervening-regular condition.

Excerpt of the tone series presented in Experiment 2 during passive listening with regular arrangement of the intervening sequence. This condition is identical to the passive overlapping-130 condition of Experiment 1 (Audio File S3).

### Audio File S10 | Experiment 2, passive listening, intervening-random condition.

As Audio File S9, but with random arrangement of the intervening sequence.

### Audio File S11 | Experiment 2, active listening, intervening-regular condition.

Excerpt of the tone series presented in Experiment 2 during active listening with regular arrangement of the intervening sequence. In contrast to the passive listening version, the intervening sequence is initially absent and then fades in with a gradual intensity increase to facilitate hearing out the test sequence. This condition is identical to the active overlapping-130 condition of Experiment 1 (Audio File S7).

### Audio File S12 | Experiment 2, active listening, intervening-random condition.

As Audio File S11, but with random arrangement of the intervening sequence.

## REFERENCES

- Andreou, L.-V., Kashino, M., and Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hear. Res.* 280, 228–235.
- Bee, M. A., and Klump, G. M. (2005). Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain. Behav. Evol.* 66, 197–214.
- Bendixen, A., Böhm, T. M., Szalárdy, O., Mill, R., Denham, S. L., and Winkler, I. (in press). Different roles of similarity and predictability in auditory stream segregation. *Learn. Percept.*
- Bendixen, A., Denham, S. L., Gyimesi, K., and Winkler, I. (2010). Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* 128, 3658–3666.
- Bendixen, A., Prinz, W., Horváth, J., Trujillo-Barreto, N. J., and Schröger, E. (2008). Rapid extraction of auditory feature contingencies. *Neuroimage* 41, 1111–1119.
- Bendixen, A., Roeber, U., and Schröger, E. (2007). Regularity extraction and application in dynamic auditory stimulus sequences. *J. Cogn. Neurosci.* 19, 1664–1677.
- Bendixen, A., and Schröger, E. (2008). Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biol. Psychol.* 78, 231–241.
- Bregman, A. S. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound.* Cambridge, MA: MIT Press.
- Bregman, A. S., Ahad, P. A., Crum, P. A., and O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Percept. Psychophys.* 62, 626–636.
- Conway, C. M., Pisoni, D. B., and Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: the auditory scaffolding hypothesis. *Curr. Dir. Psychol. Sci.* 18, 275–279.
- Costa-Faidella, J., Grimm, S., Slabu, L., Diaz-Santaella, F., and Escera, C. (2011). Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. *Psychophysiology* 48, 774–783.
- Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1119–1130.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.
- Denham, S. L., and Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* 100, 154–170.
- Dowling, W. J., Lung, K. M.-T., and Herbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Percept. Psychophys.* 41, 642–656.
- Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., and Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *Neuroimage* 54, 2267–2277.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* New York: Wiley.
- Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24, 95–112.
- Grimm, S., Widmann, A., and Schröger, E. (2004). Differential processing of duration changes within short and long sounds in humans. *Neurosci. Lett.* 356, 83–86.
- Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelić, J. H., and Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *J. Neurosci.* 25, 10494–10501.
- Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., and Winkler, I. (2008a). MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. *Psychophysiology* 45, 60–69.
- Horváth, J., Winkler, I., and Bendixen, A. (2008b). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79, 139–147.
- Horváth, J., Czigler, I., Sussman, E., and Winkler, I. (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Brain Res. Cogn. Brain Res.* 12, 131–144.
- Horváth, J., Roeber, U., and Schröger, E. (2009). The utility of brief, spectrally rich, dynamic sounds in the passive oddball paradigm. *Neurosci. Lett.* 461, 262–265.
- Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol. Rev.* 83, 323–355.
- Jones, M. R., and Boltz, M. (1989). Dynamic attending and responses to time. *Psychol. Rev.* 96, 459–491.
- Jones, M. R., Kidd, G., and Wetzell, R. (1981). Evidence for rhythmic attention. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 1059–1073.
- Kaernbach, C., and Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J. Acoust. Soc. Am.* 104, 2298–2306.
- Kujala, T., Tervaniemi, M., and Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol. Psychol.* 74, 1–19.
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide.* Mahwah, NJ: Erlbaum.
- May, P. J. C., and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47, 66–122.
- Moore, B. C. J., and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. United Acust.* 88, 320–333.
- Müller, D., Widmann, A., and Schröger, E. (2005). Auditory streaming affects the processing of successive deviant and standard sounds. *Psychophysiology* 42, 668–676.
- Näätänen, R., Astikainen, P., Ruusuvirta, T., and Huotilainen, M. (2010). Automatic auditory intelligence: an expression of the sensory-cognitive core of cognitive processes. *Brain Res. Rev.* 64, 123–136.
- Näätänen, R., Gaillard, A. W. K., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol. (Amst.)* 42, 313–329.
- Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). “Primitive intelligence” in the auditory cortex. *Trends Neurosci.* 24, 283–288.
- Neuhoff, J. G. (2003). Pitch variation is unnecessary and sometimes insufficient for the formation of auditory objects. *Cognition* 87, 219–224.
- Nousak, J. M. K., Deacon, D., Ritter, W., and Vaughan, H. G. Jr. (1996). Storage of information in transient auditory memory. *Brain Res. Cogn. Brain Res.* 4, 305–317.
- Novak, G. P., Ritter, W., Vaughan, H. G., and Wiznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalogr. Clin. Neurophysiol.* 75, 255–275.
- Paavilainen, P., Arajärvi, P., and Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *Neuroreport* 18, 159–163.
- Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., and Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology* 38, 359–365.
- Pacton, S., and Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 80–96.
- Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128.
- Rahne, T., Böckmann, M., Von Specht, H., and Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Res.* 1144, 127–135.
- Rahne, T., and Böckmann-Barthel, M. (2009). Visual cues release the temporal coherence of auditory objects in auditory scene analysis. *Brain Res.* 1300, 125–134.
- Rahne, T., and Sussman, E. (2009). Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. *Eur. J. Neurosci.* 29, 205–211.
- Remillard, G. (2009). Pure perceptual-based sequence learning: a role for visuospatial attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 528–541.
- Rimmele, J. M., Schröger, E., and Bendixen, A. (in press). Age-related changes in the use of regular patterns for stream segregation. *Hear. Res.*
- Ritter, W., Sussman, E., and Mollholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *Neuroreport* 11, 61–63.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am.* 112, 2074–2085.
- Saari, J., Paavilainen, P., Schröger, E., Tervaniemi, M., and Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *Neuroreport* 3, 1149–1151.
- Sams, M., Alho, K., and Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch

- negativity of the ERP. *Psychophysiology* 21, 434–441.
- Schröger, E. (2005). The mismatch negativity as a tool to study auditory processing. *Acta Acust. United Acust.* 91, 490–501.
- Schröger, E. (2007). Mismatch negativity: a microphone into auditory memory. *J. Psychophysiol.* 21, 138–146.
- Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., and Kaneko, S. (2000). Mismatch negativity (MMN) reveals sound grouping in the human brain. *Neuroreport* 11, 1597–1601.
- Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50.
- Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149.
- Sussman, E., Ritter, W., and Vaughan, H. G. Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Res.* 789, 130–138.
- Sussman, E., Ritter, W., and Vaughan, H. G. Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36, 22–34.
- Sussman, E., and Steinschneider, M. (2006). Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Res.* 1075, 165–174.
- Sussman, E., and Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia* 47, 771–785.
- Sussman, E., and Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Brain Res. Cogn. Brain Res.* 12, 431–439.
- Sussman, E., Čeponiene, R., Shestakova, A., Näätänen, R., and Winkler, I. (2001). Auditory stream segregation processes operate similarly in school-aged children and adults. *Hear. Res.* 153, 108–114.
- Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *J. Acoust. Soc. Am.* 117, 1285–1298.
- Sussman, E. S., Horváth, J., Winkler, I., and Orr, M. (2007). The role of attention in the formation of auditory streams. *Percept. Psychophys.* 69, 136–152.
- Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., and Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Brain Res. Cogn. Brain Res.* 25, 169–179.
- Timm, J., Weise, A., Grimm, S., and Schröger, E. (2011). An asymmetry in the automatic detection of the presence or absence of a frequency modulation within a tone: a mismatch negativity study. *Front. Psychol.* 2:189. doi:10.3389/fpsyg.2011.00189
- van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences*. Doctoral dissertation, Technical University Eindhoven, Eindhoven.
- van Zuijlen, T. L., Simoens, V. L., Paavilainen, P., Näätänen, R., and Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *J. Cogn. Neurosci.* 18, 1292–1303.
- Vliegen, J., and Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *J. Acoust. Soc. Am.* 105, 339–346.
- Walker, L. J., Carpenter, M., Downs, C. R., Cranford, J. L., Stuart, A., and Pravica, D. (2001). Possible neuronal refractory or recovery artifacts associated with recording the mismatch negativity response. *J. Am. Acad. Audiol.* 12, 348–356.
- Weise, A., Grimm, S., Müller, D., and Schröger, E. (2010). A temporal constraint for automatic deviance detection and object formation: a mismatch negativity study. *Brain Res.* 1331, 88–95.
- Weise, A., Müller, D., Grimm, S., Rüb-samen, R., and Schröger, E. (2007). Differential processing of terminal tone parts within structured and non-structured tones. *Neurosci. Lett.* 421, 163–167.
- Winkler, I. (1996). “Necessary and sufficient conditions for the elicitation of the mismatch negativity,” in *Recent Advances in Event-Related Brain Potentials Research. Proceedings of the XIth International Conference on Event-Related Brain Potentials (EPIC), Okinawa, Japan, June 25–30, 1995*, eds C. Ogura, Y. Koga, and M. Shimokochi (Amsterdam: Elsevier), 36–43.
- Winkler, I. (2003). “Change detection in complex auditory environment: Beyond the oddball paradigm,” in *Detection of Change: Event-Related Potential and fMRI Findings*, ed. J. Polich (Boston: Kluwer Academic Publishers), 61–81.
- Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163.
- Winkler, I., and Czigler, I. (1998). Mismatch negativity: deviance detection or the maintenance of the “standard”. *Neuroreport* 9, 3809–3813.
- Winkler, I., Czigler, I., Jaramillo, M., Paavilainen, P., and Näätänen, R. (1998). Temporal constraints of auditory event synthesis: evidence from ERPs. *Neuroreport* 9, 495–499.
- Winkler, I., Denham, S. L., Mill, R., Böhm, T. M., and Bendixen, A. (2012). Multistability in auditory stream segregation: a predictive coding view. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1001–1012.
- Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* 13, 532–540.
- Winkler, I., Kushnerenko, E., Horváth, J., Čeponiene, R., Fellman, V., Huotilainen, M., Näätänen, R., and Sussman, E. (2003a). Newborn infants can organize the auditory world. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11812–11815.
- Winkler, I., Sussman, E., Tervaniemi, M., Horváth, J., Ritter, W., and Näätänen, R. (2003b). Preattentive auditory context effects. *Cogn. Affect. Behav. Neurosci.* 3, 57–77.
- Yabe, H., Tervaniemi, M., Reinikainen, K., and Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport* 8, 1971–1974.
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., and Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology* 35, 615–619.
- Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., and Schecker, M. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *Neuroreport* 16, 2015–2019.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 November 2011; paper pending published: 22 January 2012; accepted: 22 April 2012; published online: 21 May 2012.

Citation: Bendixen A, Schröger E, Ritter W and Winkler I (2012) Regularity extraction from non-adjacent sounds. *Front. Psychology* 3:143. doi: 10.3389/fpsyg.2012.00143

This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*.

Copyright © 2012 Bendixen, Schröger, Ritter and Winkler. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.