Title: Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: Memory basis for predictive sound processing

Abbreviated title: Transitional probabilities in deviance detection

Authors: Maria Mittag[1,2], Rika Takegata[2], István Winkler [3]

[1] Institute for Learning & Brain Sciences, University of Washington, P.O. Box 357988 Seattle, United States of America

[2] Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 00014 Helsinki, Finland

[3] Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1519 Budapest, P.O. Box 286, Hungary

Corresponding author: István Winkler, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1519 Budapest, P.O. Box 286, Hungary USA, phone: +36-1 354 2296, email: winkler.istvan@ttk.mta.hu

ABSTRACT

Representations encoding the probabilities of auditory events do not directly support predictive processing. In contrast, information about the probability with which a given sound follows another (transitional probability) allows predictions of upcoming sounds. We tested whether behavioral and cortical auditory deviance detection (the latter indexed by the mismatch negativity event-related potential) relies on probabilities of sound patterns or on transitional probabilities. We presented healthy adult volunteers with three types of rare tone-triplets among frequent standard triplets of High-Low-High (HLH) or LHL pitch structure: proximity deviant (HHH/LLL), reversal deviant (LHL/HLH), and first-tone deviant (LLH/HHL). If deviance detection was based on pattern probability, reversal and first-tone deviants should be detected with similar latency because both differ from the standard at the first pattern position. If deviance detection was based on transitional probabilities, then reversal deviants should be the most difficult to detect, because, unlike the other two deviants, they contain no low-probability pitch transitions. The data clearly showed that both behavioral and cortical auditory deviance detection utilizes transitional probabilities. Thus the memory traces underlying cortical deviance detection may provide a link between stimulus-probability based change/novelty detectors operating at lower levels of the auditory system and higher auditory cognitive functions that involve predictive processing.

INTRODUCTION

Following on Helmholtz' idea of unconscious inferences (1860/1962), the notion of predictive information processing has become a dominant theory of perception (Gregory, 1980; Friston, 2005; Bar, 2007). However, while there is neurophysiological evidence showing error signals resulting from failed predictions (Wang et al., 2006; Alink et al., 2010), little is known about how memory representations support predictive processing. Auditory deviance detection, as reflected by the mismatch negativity (MMN) event-related brain potential (ERP) allows one to study this issue, because MMN is a prime candidate for an auditory prediction error signal (Winkler and Czigler, 2012). Early descriptions of MMN suggested that it is elicited by low-probability sounds (deviant) encountered within the context of a frequent sound (standard). However, representing stimulus probabilities does not support accurate predictions. These require knowledge of the probabilities by which sounds follow each other (transitional probabilities). Winkler (2007) suggested that the memory underlying MMN stores inter-sound relationships (transitional probabilities). MMN elicitation by temporal violations (e.g., Nordby et al., 1988) suggests that the underlying memory representations also include the expected timing of upcoming sounds (cf. Dehaene et al., 2015). Here we test the question of whether MMN generation relies on transitional probabilities or the probabilities of stimulus events (individual tones or repeating tonal patterns).

Specifically, we address the question whether transitional probabilities supersede event probabilities in cortical auditory deviance detection. At lower levels of the auditory system, stimulus probability predominantly governs change/novelty detection, such as stimulus-specific adaptation (SSA; Ulanovsky, Las & Nelken, 2003; Malmierca et al., 2014). Thus MMN could link primitive change/novelty detection mechanisms with predictive processing based cognitive functions.

We recorded ERP responses to three types of deviant tone-triplets presented among standard triplets, which alternated two tones with different pitches (e.g., low-high-low: L-H-L; Figure

3

1A). The deviant triplets were either (1) proximity (L-L-L), (2) reversal (H-L-H), or (3) first-tone deviants (H-H-L). No MMN can be expected for either of the deviants if MMN was elicited by low-probability tones, because none of the deviants include sounds that would not also appear with at least .33 probability in the standard triplet. However, previous studies showed that deviance detection is based on low-probability patterns rather than individual stimuli when the repeated pattern is short (<500 ms; Sussman et al., 1998; Sussman et al., 2002). If deviance detection were based on low-probability patterns, then the first-tone and the reversal deviant should be detected with the same timing, as both begin with a different tone than the standard. Further, the reversal deviant should produce the best behavioral detection, because it differs from the standard in all three positions, whereas the other two deviants differ in only one position. If, however, deviance detection were based on transitional probabilities, then the low-probability pitch-repetitions of the proximity and the first-tone deviant should be easy to detect, whereas detecting the reversal deviants should be difficult, because it only differs from the standard by the order between two, otherwise frequent, pitch transitions. Another possibility is that high and low tones are segregated to separate streams (Bregman, 1990). In this case, reversal and first-tone deviants produce an omission in one stream and a tone arriving "too early" in the other stream resulting in similar MMN responses. (Note that this alternative is compatible both with the event- and the transitional-probability based account of MMN.) Another similar possibility has been brought up by Deutsch (1974, 1975), who suggested that pitch proximity results in "reorganizing" perception by pitch, e.g., overruling the identity of source location. In contrast, MMN studies suggested that the auditory system forms separate memory traces for sounds presented to the two ears (Praamstra and Stegeman, 1992; Shalgi and Deouell, 2007). For contrasting these two possibilities, we also administered a dichotic condition in which the two ears always received the opposite tone pattern (Figure 1A). With pattern reorganization, the proximity

deviant should be the most difficult to detect in the dichotic condition, as it matches the reorganized version of the standard pattern.

MATERIALS AND METHODS

Participants

Fourteen right-handed healthy adults (mean age: 25.9 years, SD = 6.1 years, four males) with normal hearing participated in the experiment. Data from one participant was discarded from the behavioral analysis of the experiment due to low hit rates (<50%). The mean age of the remaining thirteen participants for behavioral analysis was 26.2 years (SD = 6.3 years, three males). Participants gave written informed consent after the experimental procedure was explained to them. They received modest payment for their participation. The study was conducted in full accordance with the Declaration of Helsinki and all applicable national laws, and it was approved by the Ethical Committee of the Department of Psychology, University of Helsinki.

>> Insert Figure 1 about here >>

Stimulus material

Stimuli were triplets comprising low and high sinusoidal tones (784 and 988 Hz, respectively; 75-ms tone duration including 10 ms linear rise and fall times; 50 dB intensity above the participant's hearing threshold measured with a staircase procedure using the same tones). Tones were created by Cool Edit 2000 (Ellison and Johnston, 2000). They were presented with a stimulus onset asynchrony (SOA; onset-to-onset interval) of 135 ms and an inter-pattern (offset-to-onset) interval of 200 ms. The SOA and the inter-pattern interval were selected to allow automatic grouping of the tone triplets, which for this type of stimuli requires short (<500 ms) pattern duration (Sussman et al., 1998; Sussman et al., 2002). In the

binaural condition, the same pattern was presented to both ears. In the dichotic condition, the two ears always simultaneously received different tones.

In the dichotic condition (Figure 1A, right panel), the frequent (standard; p = .875) triplet was the low-high-low (L-H-L) pattern presented to the left and the high-low-high (H-L-H) pattern presented to the right ear. Three types of rare (deviant; p = .125) triplets were presented in separate stimulus blocks: (1) the proximity deviant had the frequency of the second tone reversed compared to the standard triplet (L-L-L/H-H-H to the left and right ears, respectively); (2) the reversal deviant had the frequencies of all three tones reversed compared to the standard pattern (H-L-H/L-H-L to the left and right ears, respectively); and (3) the first-tone deviant had the frequency of the first tone reversed (H-H-L/L-L-H to the left and right ears, respectively). Each of the three sequences (differing in the deviant triplet) received two stimulus blocks. In the binaural condition (Figure 1A, left panel), identical tones were delivered to both ears. The standard triplet was, in separate stimulus blocks, either the left- or the right-ear standard triplet of the dichotic condition (L-H-L or H-L-H, respectively) and, again in separate stimulus blocks, the deviants matched the pattern presented to the corresponding ear in the dichotic condition. Thus, the structure of the stimulus conditions was Stimulation [dichotic vs. binaural] × Deviant Type [proximity, reversal, first tone] with each binaural condition receiving two different stimulus blocks (collapsed in all analyses) and each dichotic condition two identical ones (altogether 12 stimulus blocks).

The order of the tones within each sequence was pseudorandomized with the limitation that deviant triplets were separated by at least three standard triplets and the first five triplets of each sequence included only standards. Stimuli were delivered by Presentation software (version 9.90 Neurobehavioral Systems, Davis, CA) via MDR-7506 (Sony, Tokyo, Japan) headphones.

Design and procedure

6

Participants were seated in a comfortable armchair in a sound-attenuated and electrically shielded room of the Cognitive Brain Research Unit of the University of Helsinki. The experiment consisted of two halves. In the passive condition, which was always administered first, participants were instructed to watch a silent movie of their choice, and to ignore the sounds. In the active condition, participants were instructed to press a button to the rare deviant triplets as quickly as possible and without sacrificing accuracy. Separately, in each half, the order of stimulus blocks was counterbalanced both within and across participants. Stimulus blocks in the passive condition consisted of 85 deviants and 425 standards (in two stimulus blocks, 170 deviants and 850 standards for each condition), whereas stimulus blocks of the active conditions consisted of 40 deviants and 200 standards (in two stimulus blocks, 80 deviants and 400 standards for each condition). Note that only the behavioral data is reported from the active condition.

Analysis of the behavioral data

Correct responses (hits) were defined as button presses occurring 100-1000 ms after deviance onset. This was different across the different deviants, because the proximity deviant started to differ from the standard by the 2nd tone of the triplet, whereas the other two deviants already differed from the standard by 1st tone of the triplet. Only correct responses were included in the analysis of reaction times (RTs), which were measured from the onset of the deviations. False alarms (FAs) were defined as button presses outside of the time window for correct responses, possibly indicating an incorrect identification of the standard as a deviant. Grier's A' sensitivity index was used for assessing discrimination sensitivity (Grier, 1971). Behavioral measures were statistically analyzed with two-way repeated-measures ANOVAs with Stimulation [dichotic vs. binaural] X Deviant Type [proximity, reversal, first-tone] as factors. For all ANOVAs, Greenhouse-Geisser correction was applied where appropriate; p values after correction, ε correction values, and partial η^2 effect sizes are reported together

with the original degrees of freedom. Bonferroni's correction was applied to post hoc analyses, where necessary.

EEG recording and preprocessing

The electroencephalogram (EEG) was recorded with a 28-channel electrode cap from locations evenly covering frontal, central, temporal, and parietal areas of the scalp and from two electrodes placed at the left and right mastoids (DC-40 Hz band pass, sampling rate 500 Hz, NeuroScan Synamp[2] amplifier; Compumedics Ltd.). The tip of the nose served as the common reference electrode. Signals were online referenced to the average of all electrodes, then offline re-referenced to the nose lead. Vertical and horizontal eye movements were monitored by bipolarly recording the electrooculogram (EOG) from two pairs of electrodes, one pair placed above and below the right eye and the other attached lateral to the outer canthi of the eyes. All inter-electrode impedances were set below 10 kΩ.

ERP analysis

The EEG was off-line band pass-filtered (1-20 Hz, 24 Hz/octave). Epochs of 550 ms duration including a 100-ms pre-stimulus period (serving as baseline for the amplitude measurements) were extracted from the continuous EEG, and separately averaged for each condition and stimulus type. For the reversal and first-tone deviants, the pre-stimulus period ended at the onset of the first tone, whereas for the proximity deviant, the pre-stimulus period ended at the onset of the second tone. Thus the epochs for each deviant and their corresponding standard triplet were anchored at the point where the standard and the deviant started to differ from each other (treated as the 0 ms point of the ERP responses). Epochs with the EEG or EOG amplitude exceeding 100 µV at any electrode were automatically rejected. MMN responses were assessed by subtracting the standard-stimulus ERP from the corresponding deviant stimulus ERP, separately for each condition and deviant. After artifact rejection, on average

130 accepted deviant-stimulus trials (range: 124-164) were analyzed in the passive condition and 69 accepted deviant-stimulus trials (range: 65-77) in the active condition.

Based on visual inspection of the traces, individual deviance-response peak latencies were measured from the difference waveforms as the most negative frontal (Fz) peaks within a window of 150-350 ms post-deviance-onset for the proximity deviant, 350-550 ms for the reversal deviant, and 200-400 ms for the first-tone deviant, uniformly in the binaural and the dichotic condition. Because previous studies have shown that the MMN latency can be considerably delayed for pattern deviants (e.g., Winkler and Schröger, 1995), these rather late deviance-related responses were regarded as MMNs (the issue of the long MMN peak latency for reversal deviants is discussed in detail in Discussion). A two-way repeated-measures ANOVA was conducted on the peak latencies for testing the effects of Stimulation [dichotic vs. binaural] and Deviant Type [proximity, reversal, first-tone].

MMN amplitudes were measured as the mean voltage within 30-ms time windows centered on the peak in the group-average difference waveform. One-tailed t-tests were used to determine whether the frontal (Fz) MMN amplitudes significantly differed from zero. A two-way repeated-measures ANOVA was conducted for testing the effects of Stimulation [dichotic vs. binaural] and Deviant Type [proximity, reversal, first tone] on the MMN amplitudes measured from a fronto-central region of interest (ROI). This ROI was selected for the tests because the MMN has its maximum at fronto-central sites (for recent reviews, see Näätänen et al., 2005; Kujala et al., 2007). The ROI was calculated by averaging the amplitudes over the following electrode locations: FP1, FP2, F3, Fz, F4, FC5, FC1, FC2, FC6, C3, Cz, C4 (shown in Figure 1B).

For all ANOVAs, Greenhouse-Geisser correction was applied where appropriate; p values after correction, ε correction values, and partial η^2 effect sizes are reported together with the original degrees of freedom. Bonferroni's correction was carried out for all post hoc analyses.

RESULTS

Behavioral data

A significant interaction between Stimulation and Deviant Type was found for hit rates $(F(2,24) = 15.5, p < .001, \eta^2 = .563, \varepsilon = .888)$, discrimination sensitivity $(F(2,24) = 10.57, p = .001, \eta^2 = .468, \varepsilon = .804)$, and reaction times $F(2,24) = 8.49, p = .002, \eta^2 = .414, \varepsilon = .973)$. These interactions were followed up by exploring the effects of the Deviant type, separately for the dichotic and the binaural condition.

In the binaural but not in the dichotic condition, participants responded significantly less accurately to reversal (73%) than to proximity (96%, p = .001) and first-tone deviants (93%, p = .003), and discrimination sensitivity was significantly lower for reversal (0.929) than for proximity (0.987, p = .002) and first-tone deviants (0.979, p = .011). In both conditions, RTs to reversal deviants were longer than to proximity (dichotic condition: 534 ms vs. 379 ms, p <.001; binaural condition: 612 ms vs. 397 ms, p < .001) and first-tone deviants (dichotic condition: 443 ms, p < .001; binaural condition: 482 ms, p < .001), the differences being larger in the binaural than in the dichotic condition (p ≤ .001 - .025). Further, RTs to first-tone deviants were longer than to the proximity deviants in both conditions (p < .001, both).

In sum, the reversal deviant was more difficult to detect (lower hit rates and discrimination sensitivity and longer RTs) than the other two deviants, and more so in the binaural than in the dichotic condition.

>> Insert Figure 2 at about here >>

ERP data

MMN amplitudes

Figure 2A shows the group-average frontal (Fz) ERP responses elicited by the standard and deviant triplets. The scalp-distribution maps of the difference waveforms taken from 30-ms

time windows centered at the MMN peaks (Figure 2B) are compatible with the well-known scalp distribution of the MMN response. The frontal (Fz) MMN amplitudes were significantly different from zero for each deviant and stimulation type (Table 1; see Figure 3 for scatter plots representing the full distributions of the amplitude measures). A significant interaction between Stimulation and Deviant Type was found for ROI MMN amplitude measure ($F(2,26)$ = 6.8, p = .008, η^2 = .343, ε = .819). The interaction was followed up by exploring the effects of Deviant type, separately for the two stimulation conditions.

Significantly smaller MMN amplitude was elicited by the reversal deviant than by the proximity deviant in both conditions (dichotic: p = .039; binaural: p = .006), whereas in the dichotic but not in the binaural condition, the amplitude of the reversal-deviant MMN was also significantly lower than that elicited by the first-tone deviant (p = .004) (Figure 2C). To assess the possible biasing effect of high-pass filtering on the amplitudes of the long-latency components, the statistical analysis was repeated with the reversal-deviant MMN amplitude measured with respect to the average voltage in the 100-ms interval preceding the onset of the third tone of the triplet and the first-tone deviant with respect to the average voltage in the 100-ms interval preceding the second tone. These measurement alternatives are based on the assumption that the reversal-deviant MMN is triggered by the transition between the second and the third tone and the first-tone deviant by the transition between the first and the second tone of the respective deviant triplets (see Discussion). The results remained very similar with the exception that in the dichotic condition, the MMN amplitude did not significantly differ between the reversal and the proximity deviant.

In sum, the reversal deviant elicited lower-amplitude MMN than the other deviants and the first-tone deviant elicited higher-amplitude MMN in the dichotic than in the binaural condition.

>> Insert Table 1 at about here >>

Peak latencies

A significant interaction between Stimulation and Deviant Type was found for the MMN peak latencies ($F(2,26) = 8.96$, $p = .003$, $\eta^2 = .408$, $\varepsilon = .747$). The MMN peak latency for the proximity deviant was significantly shorter in the dichotic than in the binaural condition ($p = 0.001$; see Table 1) whereas the similar comparisons for the other two deviants did not yield significant effects. There was also a significant main effect of Deviant Type ($F(2,26) = 224.5$, $p < .001$, $\eta^2 = .945$, $\varepsilon = .919$): the reversal-deviant MMN peak latencies were significantly longer than those elicited by the proximity ($p < .001$) and the first-tone deviants ($p < .001$) and the peak latencies for the proximity deviants were significantly shorter than those for the first-tone deviants ($p < .001$).

In sum, the peak latency of the deviance-related response to the reversal deviants was longer than those for the other two deviants and proximity-deviants elicited earlier MMNs in the dichotic than in the binaural condition.

DISCUSSION

We found lower hit rates and discrimination sensitivity together with longer RTs in response to the reversal than to the proximity or the first-tone deviant. Consistent with the behavioral data, smaller and later MMN responses were obtained for the reversal than for the other two deviants. This pattern of data matches the prediction based on the hypothesis that auditory deviance detection is based on detecting low-probability transitions between successive sounds. Unlike the other two deviants, the reversal deviant did not include rare pitch transitions. Therefore, this deviant was more difficult to detect than the other two. The >140 ms difference between the MMN peak of the reversal and the other two deviants is compatible with the notion that MMN was elicited as a result of deviance at the first pitch

12

transition for proximity and first-tone deviants, whereas for the reversal deviant, deviance detection could occur either on the first or the second pitch transition. If MMN to the reversal deviant was elicited by the first pitch transition, then it was based on the rare combination of position and pitch transition. Deviants differing from the standard in the combination of two features typically elicit late (>200 ms) MMN responses (see, e.g., Winkler, 2005; Winkler et al., 2005). On the other hand, it is also possible that the reversal-deviant MMN was only triggered by the second pitch transition (i.e., by the third tone), at which point the order of pitch transitions was violated.

The elicitation of MMN by all three types of deviants is incompatible with the prediction based on the assumption that the ERP responses of auditory deviance detection are triggered by low-probability sounds or by stronger adaptation for frequent than for infrequent sounds (May and Tiitinen, 2009), because then one would expect the sound with the lower probability (.33) to also elicit MMN or a less adaptive response within the standard triplets. The prediction drawn on the hypothesis that deviance detection is based on detecting low-probability patterns also does not match the current data: reversal-deviants were detected with longer RTs and elicited later MMN responses than the first-tone deviant and they were also the most difficult to detect amongst the three deviant types.

Predictions drawn on the assumption that the high and low tones were segregated or reorganized by pitch proximity are also contrasted by the observed pattern of the data. If the two sets of tones were segregated, then the first tone of both the reversal and the first-tone deviant produced omission of a tone in one stream and inclusion of a tone in the other stream. Inclusions (stimuli presented too early within an otherwise isochronous sequence) elicit MMN (e.g., Hari et al., 1989; Nordby et al., 1988). Therefore, if the streams were segregated, one should expect tone inclusions to elicit MMN. However, we found no MMN to reversal and first-tone deviants in the expected time range (100-150 ms from the onset of first tone of these deviants). This suggests that the tone sequences were not or only very seldom

13

segregated. The current stimulus sequence was almost identical to that of van Noorden (1975), who has assessed the effects of frequency separation and presentation rate on the segregation of high and low tones. Stimulus parameters place our sequences in the "ambiguous region", for which listeners can voluntarily control whether they perceive a series of tone triplets or separate high and low streams. The lack of MMN to the rare tone inclusions produced by first-tone and reversal deviants suggests that outside the focus of attention, such sequences are not segregated into two streams.

The pattern of data was generally similar between the binaural and the dichotic condition. This contrasts the prediction based on the hypothesis that pitch proximity overrules the identity of source location in grouping sounds (Deutsch, 1974, 1975; for an interpretation, see Kubovy and Van Valkenburg, 2001), because then the proximity-deviant should have been the most difficult one to detect, as it matched the percept "reorganized" on the basis of pitch proximity. However, the proximity-deviant proved to be the easiest to detect and it elicited MMN with the shortest peak latency of the three deviants. Thus it appears likely that separate representations were formed for the tonal patterns delivered to the two ears (see Praamstra and Stegeman, 1992; Shalgi and Deouell, 2007). As was noted above, the current stimulus parameters did not strongly promote segregating the tones by pitch (van Noorden, 1975). Since our tonal patterns were quite short (unlike the stimulus sequences of Deutsch, 1974 and 1975), grouping/segregation by pitch had no time to develop and to override the default organization by location. This suggests that, at least initially, the auditory system establishes transitional probabilities separately within each ear as opposed to grouping them by pitch proximity.

Thus, the observed pattern of data suggests that cortical deviance detection uses transitional probabilities rather than the probability of individual sounds or patterns. This conclusion is compatible with previous observations of MMN being elicited by "local rules" (i.e., rules allowing to predict from a sound to the next; Horváth et al., 2001; Paavilainen et al., 2007;

Bendixen et al., 2008). Further, it is also compatible with findings showing that melodic contour and even pitch intervals are retained in memory (Edworthy, 1985; Peretz and Babaï, 1992) and utilized by functions, such as deviance detection (e.g., Saarinen et al., 1992; Tervaniemi et al., 2006). That is, the auditory system is set to look for transitional probabilities within sequences (Wacongne et al., 2011; Wacongne et al., 2012), supporting the notion that the auditory system is intrinsically predictive (Friston, 2005; Bendixen et al., 2012; Winkler and Czigler, 2012; Khouri and Nelken, 2015). Indeed, there is mounting evidence that the auditory system prepares for predictable sounds (e.g., Bendixen et al., 2009; Barascud et al., 2016; Koelsch et al., 2016).

The current results also demonstrate the relationship between "transition and timing" analysis and "chunking", two lower-level processes of the brain's analysis of sequences (Deheane et al., 2015). Temporal grouping (chunking) occurs even when the sounds are task-irrelevant, provided that the cycle is sufficiently short (<500 ms; Sussman et al., 1998, 2002). This confounds the interpretation of the MMN elicited by tone repetitions in a sequence of two alternating tones (ABAB…; e.g., Horváth et al., 2001): The elicitation of MMN by tone repetitions can be interpreted in terms of transitional probabilities (i.e., tone repetition being a rare transition) or as the result of the sequence being processed in terms of repeating tone pairs (i.e., in this case, AB is the frequent standard and AA is the rare deviant pair – for evidence showing that temporal grouping may occur in such sequences, see, Brochard et al., 2003; Potter et al., 2009; Bouwer and Honing, 2015). With short–duration repeating tone patterns, chunking overtakes stimulus-based processing: in a sequence of AAAABAAAAB… structure and an SOA of 100 ms, the B tone did not elicit MMN, even though the same tones elicited MMN when the A and B tones were presented in a random order (80% and 20% for the A and B tones, respectively; Sussman et al., 1998). With longer cycles, ERP responses reflecting chunk-based processing of the sequences have only been obtained when participants attended the sounds (Sussman et al., 2002; Bekinschtein et al., 2009). The current

results provided new information by revealing that chunking is based on transitional probabilities between individual sounds, because the current sequences were processed in terms of tonal groups, which were represented by the order of transitions between sounds (see the discussion of these conclusions above).

We also found that the proximity deviant elicited earlier deviance-related responses and shorter RTs than the first-tone deviant. This is possibly due to the generation of a more precise prediction of the second tone when the first tone fully matches the standard (proximity deviant) than when it differs from it (first-tone deviant). This is supported by Barascud et al's (2016) finding showing that when the initial segment of a sound pattern matches the beginning of the preceding pattern, later deviations from the first pattern are detected. Finally, proximity deviants elicited only a single MMN response, even though they included two successive deviant pitch transitions. Previous studies showed that when two deviations deterministically follow each other within 200 ms, as was the case for the current proximity deviant, only the first one elicits the MMN response (e.g., Sussman et al., 1999; Sussman and Winkler, 2001). This phenomenon shows that, at least within a short period of time, the auditory system treats two yoked deviations as a single one.

In summary, we found strong evidence for the auditory system representing and utilizing transitional probabilities over probabilities of individual sensory events. Forming representations for transitional probabilities allows predictions of upcoming sounds. Therefore, the memory traces underlying cortical deviance detection may provide a link between stimulus-probability based change/novelty detectors operating at lower levels of the auditory system and higher auditory cognitive functions of predictive nature.

REFERENCES

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. J Neurosci 30:2960-2966.

Bar M (2007) The proactive brain: using analogies and associations to generate predictions. Trends Cogn Sci 11:280-289.

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc Natl Acad Sci U S A 113:E616-625.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009). Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA 106:1672–1677.

Bendixen A, Schröger E, Winkler I (2009) I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. J Neurosci 29:8447-8451.

Bendixen A, SanMiguel I, Schröger E (2012) Early electrophysiological indicators for predictive processing in audition: a review. Int J Psychophysiol 83:120-131.

Bendixen A, Prinz W, Horváth J, Trujillo-Barreto NJ, Schröger E (2008) Rapid extraction of auditory feature contingencies. NeuroImage 41:1111-1119.

Bregman AS (1990) Auditory Scene Analysis. Cambridge, MA: MIT Press.

Bouwer FL, Honing H (2015). Temporal attending and prediction influence the perception of metrical rhythm: Evidence from reaction times and ERPs. Frontiers in Psychology 6:1094.

Brochard R, Abecasis D, Potter DD, Ragot R, Drake C (2003). The "ticktock" of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. Psychological Science 14(4):362–366.

Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C (2015) The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. Neuron 88(1):2-19.

Deutsch D (1974) An auditory illusion. Nature 251:307-309.

Deutsch D (1975) Two-channel listening to musical scales. J Acoust Soc Am 57:1156-1160.

Edworthy J (1985) Interval and contour in melody processing. Music Perception 2:375–388.

Ellison R, Johnston D (2000) Software Cool Edit. In. Phoenix (USA): Syntrillium Software Corporation.

Friston K (2005) A theory of cortical responses. Philosophical Transactions of the Royal Society of London B: Biol Sci 360:815-836.

Gregory RL (1980) Perceptions as hypotheses. Philosophical Transactions of the Royal Society of London B: Biol Sci 290:181-197.

Grier, JB (1971) Nonparametric indexes for sensitivity and bias: computing formulas. Psycholog Bull 75(6):424-429.

Hari R, Joutsiniemi SL, Hämäläinen M, Vilkman V (1989) Neuromagnetic responses of human auditory cortex to interruptions in a steady rhythm. Neurosc Lett 99:164-168.

Helmholtz H (1860/1962) Handbuch der Physiologischen Optik., English trans. Edition. New York: Dover.

Horváth J, Czigler I, Sussman E, 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.

Khouri L, Nelken I (2015) Detecting the unexpected. Curr Opin Neurobiol 35:142-147.

Koelsch S, Busch T, Jentschke S, Rohrmeier M (2016) Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. Sci Rep 6:19741.

Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80:97-126.

Kujala T, Tervaniemi M, Schröger E (2007) The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biol Psychol 74:1-19.

Malmierca MS, Sanchez-Vives MV, Escera C, Bendixen A (2014) Neuronal adaptation, novelty detection and regularity encoding in audition. Front Syst Neurosci 8:111.

May PJC, Tiitinen H (2009) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiol 46:1–57.

Näätänen R, Jacobsen T, Winkler I (2005) Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. Psychophysiol 42:25-32.

Nordby H, Roth WT, Pfefferbaum A (1988) Event related potentials to time deviant and pitch deviant tones. Psychophysiol 25:249-261.

Paavilainen P, Arajärvi P, Takegata R (2007) Preattentive detection of nonsalient contingencies between auditory features. NeuroReport 18:159-163.

Peretz I, Babaï M (1992) The role of contour and intervals in the recognition of melody parts: evidence from cerebral asymmetries in musicians. Neuropsychol 30:277–292.

Potter DD, Fenwick M, Abecasis D, Brochard R (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. Cortex 45(1):103–109.

Praamstra P, Stegeman DF (1992) On the possibility of independent activation of bilateral mismatch negativity (MMN) generators. Electroencephalogr Clin Neurophysiol 82:67-80.

Saarinen J, Paavilainen P, Schröger E, Tervaniemi M, Näätänen R (1992) Representation of abstract attributes of auditory stimuli in the human brain. NeuroReport 3:1149–1151.

Shalgi S, Deouell LY (2007) Direct evidence for differential roles of temporal and frontal components of auditory change detection. Neuropsychologia 45:1878-1888.

Sussman E, Ritter W, Vaughan HG, Jr. (1998) Predictability of stimulus deviance and the mismatch negativity. NeuroReport 9:4167-4170.

Sussman E, Winkler I (2001) Dynamic sensory updating in the auditory system. Cogn Brain Res 12:431-439.

Sussman E, Winkler I, Ritter W, Alho K, Näätänen R (1999) Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. Neurosc Lett 264: 161-164.

Sussman E, Winkler I, Wang W (2003) MMN and attention: competition for deviance detection. Psychophysiol 40:430-435.

Sussman E, Winkler I, Huotilainen M, Ritter W, Näätänen R (2002) Top-down effects can modify the initially stimulus-driven auditory organization. Brain Res Cogn Brain Res 13:393-405.

Tervaniemi M, Castaneda A, Knoll M, Uther M (2006) Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. NeuroReport 17:1225–1228

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391-398.

van Noorden LPAS (1975) Temporal Coherence in the Perception of Tone Sequences. In: Technical University Eindhoven.

Wacongne C, Changeux JP, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. J Neurosci 32:3665-3678.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci U S A 108:20754-20759.

Wang W, Jones HE, Andolina IM, Salt TE, Sillito AM (2006) Functional alignment of feedback effects from visual cortex to thalamus. Nat Neurosci 9:1330-1336.

Winkler I (2007) Interpreting the mismatch negativity. J Psychophysiol 21:147–163.

Winkler I, Schröger E (1995) Neural representation for the temporal structure of sound patterns. NeuroReport 6:690-694.

Winkler I, Czigler I (2012) Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. Int J Psychophysiol 83:132-143.

Winkler I, Schröger E (2015) Auditory perceptual objects as generative models: Setting the stage for communication by sound. Brain Lang 148:1-22.

Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13:532-540.

Winkler I, Czigler I, Sussman E, Horváth J, Balázs L (2005) Preattentive binding of auditory and visual stimulus features. J Cogn Neurosci 17(2):320-39.

Winkler I, Takegata R, Sussman E (2005) Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Cogn Brain Res 25:291-299.

FIGURE CAPTIONS

Figure 1. (A) Schematic illustration of the tone patterns. Stimuli were triplets composed of low (black rectangle) and high sinusoidal tones (white rectangle). The same tone pattern was presented to both ears in the binaural condition (left panel; the 2 variants are depicted separately at the top and the bottom), whereas different patterns were simultaneously presented to the left and right ears in the dichotic condition (right panel). Frequent standard triplets were interspersed with one of three rare deviant triplets: (1) proximity, (2) reversal, and (3) first-tone deviant. Tone timing is marked under the 1st variant of tone patterns presented in the binaural condition shown in the upper left corner. (B) Map of electrodes selected for the ROI analysis.

Figure 2. ERP responses in the passive conditions. (A) Group-average (N = 14) frontal (Fz) ERPs overplotted for the standard (thin blue line) and, in separate rows, the three different deviant triplets (left column, dotted red lines) together with the deviant-minus-standard difference waveforms (right column, bold black lines), separately for the passive binaural (left panel) and the passive dichotic condition (right panel). The MMN measurement intervals are marked by grey rectangles. The green shaded areas around the difference waveforms represent the 95% confidence intervals of the group mean. Calibration is at the upper left corner. A schematic illustration of the tone triplets appears below the ERP waveforms. (B) Scalp distribution maps of the difference waveforms from the 30-ms time window centered at the MMN peaks. Color calibration is on the right side. (C) Difference waveforms averaged over the frontal ROI for the three deviants (proximity: thick yellow line; reversal: thin green line; first-tone: dotted purple line), separately for the binaural (left) and the dichotic condition (right). Calibration is shown at the binaural-condition waveforms.

Figure 3. Scatter plots of the full distributions of individual mean MMN amplitudes. The binaural condition is presented on the left, the dichotic condition on the right panel: proximity deviant (blue circle), reversal deviant (green cross), first tone deviant (red diamond).
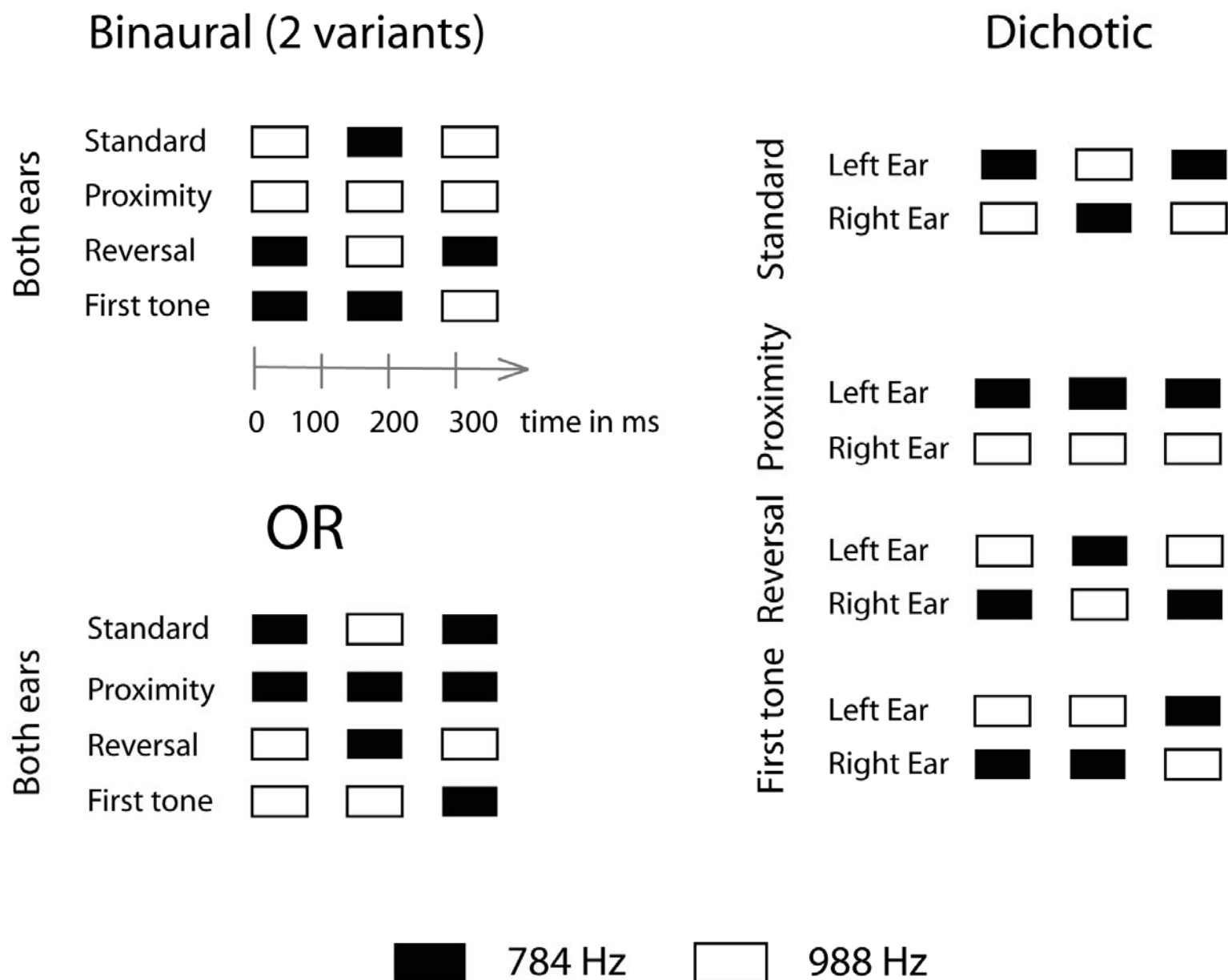
Table 1

Group-mean (N=14) frontal (Fz) MMN amplitudes (µV) and peak latencies (ms) for the Binaural and Dichotic conditions. Standard deviations are shown in parentheses.

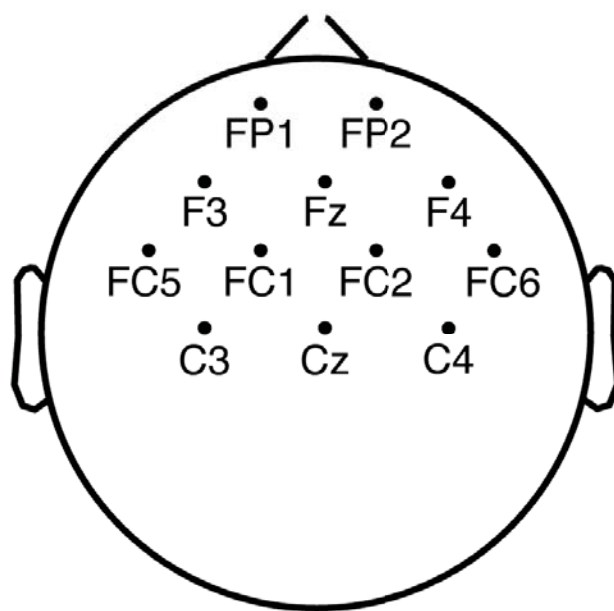| Deviant	Binaural Mean Amplitude	Peak Latency	Dichotic Mean Amplitude	Peak Latency
Proximity	-2.49 (1.7)****	238	-1.84 (1.7)***	190
Reversal	-.49 (.8)*	416	-.94 (1.2)**	450
First tone	-1.46 (1.4)***	274	-3.26 (1.9)****	300

MMN mean amplitudes differed from zero (one-tailed t-tests): *p < .05, **p < .01, ***p < .005, ****p < .001.
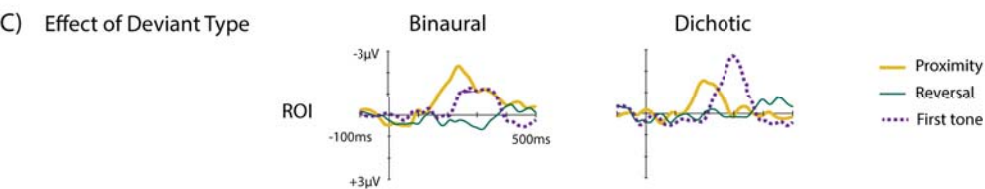Peak latencies for all deviants are reported from the deviance onset.

A) Experimental Design

Binaural (2 variants)

Both ears

	0	100	200
Standard	□	■	□
Proximity	□	□	□
Reversal	■	□	■
First tone	■	■	□

0 100 200 300   time in ms

OR

Both ears

Standard	■	□	■
Proximity	■	■	■
Reversal	□	■	□
First tone	□	□	■

Dichotic

Standard
| Left Ear | ■ | □ | ■ |
| Right Ear | □ | ■ | □ |

Proximity
| Left Ear | ■ | ■ | ■ |
| Right Ear | □ | □ | □ |

Reversal
| Left Ear | □ | ■ | □ |
| Right Ear | ■ | □ | ■ |

First tone
| Left Ear | □ | □ | ■ |
| Right Ear | ■ | ■ | □ |

■ 784 Hz   □ 988 Hz

B) Electrode map

Passive Condition (N = 14)

A)

Binaural — ERP Responses — Difference waveforms

Dichotic — ERP Responses — Difference waveforms

Proximity

Fz

-3μV

-100ms   500ms

+3μV

Reversal

First tone

Standard
Deviant
Difference waveforms
Confidence interval (95%) of group mean

B) Scalp distribution maps

Binaural   Dichotic

Proximity   Reversal   First tone   Proximity   Reversal   First tone

-3

0

3μV

C) Effect of Deviant Type

Binaural   Dichotic

ROI

-3μV

-100ms   500ms

+3μV

Proximity
Reversal
First tone

Individual mean MMN amplitudes