From Sensory to Long-Term Memory

Evidence from Auditory Memory Reactivation Studies

István Winkler[1,2] and Nelson Cowan[3]

[1] Institute for Psychology, Hungarian Academy of Sciences, Hungary,
[2] Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, Finland,
[3] Department of Psychological Sciences, University of Missouri, USA

Abstract. Everyday experience tells us that some types of auditory sensory information are retained for long periods of time. For example, we are able to recognize friends by their voice alone or identify the source of familiar noises even years after we last heard the sounds. It is thus somewhat surprising that the results of most studies of auditory sensory memory show that acoustic details, such as the pitch of a tone, fade from memory in ca. 10–15 s. One should, therefore, ask (1) what types of acoustic information can be retained for a longer term, (2) what circumstances allow or help the formation of durable memory records for acoustic details, and (3) how such memory records can be accessed. The present review discusses the results of experiments that used a model of auditory recognition, the auditory memory reactivation paradigm. Results obtained with this paradigm suggest that the brain stores features of individual sounds embedded within representations of acoustic regularities that have been detected for the sound patterns and sequences in which the sounds appeared. Thus, sounds closely linked with their auditory context are more likely to be remembered. The representations of acoustic regularities are automatically activated by matching sounds, enabling object recognition.

Keywords: memory, auditory sensory memory, long-term sensory memory, reactivation, event-related brain potentials, mismatch negativity (MMN)

1. Introduction

Traditionally, the processing of sensory and categorical information have been distinguished from each other on the basis of performance differences found in some experimental procedures and by evidence showing anatomical separation of sensory and categorical processing in the human brain (see, however, Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001, for recent evidence concerning the "intelligent" functions of auditory sensory-specific areas in the human brain). Previous research in experimental psychology identified four important features characterizing sensory memory traces, distinguishing them from categorical memory representations (Broadbent, 1958; Cowan, 1984, 1988; for a full discussion of the supporting and contradictory evidence, see Section 3.2.): 1) the formation of sensory memory traces does not depend on attention; 2) the information stored in sensory memory traces is modality-specific; and 3) has a resolution, which is finer than the conventional meaningful categories; but 4) it is lost within a short period of time. The goal of the current review is to reexamine these distinguishing features of sensory memory traces in the light of recent evidence obtained with electrophysiological and behavioral methods, mostly using memory-reactivation procedures.

1.1 Properties of Sensory Memory: Behavioral Studies

Classical multi-store models of memory postulated separate stores for the retention of sensory information (e.g., Atkinson & Shiffrin, 1968). The sensory memory stores (a separate one for each modality) were assumed to serve as temporary buffers from which information could be accessed for a short time, after which they were lost due to decay or to interference from more recent stimuli. The information selected from the sensory buffers was categorized, or transformed into a common internal code, allowing modality-independent operations. Only categorized information was assumed to be stored in more durable stores.

These features of the multi-store models correspond well with the majority of the results of sensory memory research. For example, it has been found in

many studies that subjects can only tell the difference between two closely similar sounds if the sounds to be compared are presented within ca. 10 s (taken to suggest decay) Cowan, 1984). Also, presenting irrelevant sounds between the to-be-compared ones deteriorates performance, the more so, the closer the similarity between the to-be-compared and the intervening sounds (taken to suggest interference) (Cowan, 1984; Deutsch, 1975; Massaro, 1970). When the sounds to be compared are separated by long silent intervals, subjects can only discriminate them if they substantially differ from each other. The degradation of sensory resolution is compatible with the idea that once the trace of the first sound is eliminated from the sensory buffer, subjects can only rely on those memory stores that have a longer retention interval. These stores, however, only contain categorized information representing stimulus features with a cruder resolution than the sensory buffer. Importantly, it appears that the retention interval of the sensory buffer cannot be extended much by top-down control (e.g., Keller, Cowan, & Saults, 1995), though there is a small effect of that nature. Three related findings in the area of speech perception leading toward the notion of separated sensory and categorical memory stores are (1) the category boundary effect, or more successful discrimination between two phonemes falling across a category boundary than between two allophones (instances of a single category) of comparable physical separation, (2) the delay effect, or poorer comparison of two allophones as a function of the temporal separation between these allophones in the range of a few seconds, and (3) the vowel advantage, or much more rapid forgetting of allophonic detail for the acoustically-complex stop consonants than for the acoustically-simpler vowels (Fujisaki & Kawashima, 1971; Pisoni, 1973). Acoustic theory has emphasized this distinction between unstable sensory information and more stable categorical information (e.g., Durlach & Braida, 1969).

However, there also exist results indicating that not all sensory information is lost within a few seconds. Craik and Kirsner (1974) reviewed studies showing that people often remember voice information for longer than the period of 30 s or so that has been the presumed duration of auditory sensory memory, and they carried out their own interesting experiments reinforcing that point. In their studies, spoken target words were presented 4 s apart in two different voices (male and female) in random order. The test for each word consisted of another version of the word (the probe word) that had to be recognized. Foils that had not been presented as targets also appeared as negative probes, and the question was whether the probe had

appeared before in the experiment. The distance between the target and the corresponding probe word was a lag of 1, 2, 4, 8, 16, or 32 words. At a lag of 1, the target and probe occurred with no intervening word; at a lag of 2, they occurred with one intervening word; and so on. It was found in several of the experiments that spoken probe words were recognized faster and more accurately when the voice remained the same from the target word to the probe word. This occurred even at the long delays, indicating that some memory of the voice persisted for over 2 min in the presence of intervening words. It also was found that subjects maintained an ability to recall (at better-than-chance accuracy) the voice in which a particular word had been presented. For example, in one experiment, in which the probe words were visually presented, the proportion of recall of the voice conditional upon correct recognition of the word was, at the six lags, 1.00, 0.98, 0.87, 0.75, 0.76, and 0.73, respectively. Thus, at lags of 8 words (32 s) and above, recall of the voice reached an asymptote of much-better-than-chance accuracy.

In a very different type of procedure demonstrating long-term storage of memory for sound, Crowder (1989) presented a pure tone followed by a note played by a musical instrument. Comparison of the pitches of the two sounds was speeded when participants had advanced knowledge of which instrument was to be used for the second sound. Presumably, long-term memory was used to generate a mental image of the frequency of the pure tone as played in the timbre of the instrument that was used for the second sound.

As a consequence, modern memory models suggest that sensory stimulus codes are processed along with categorical stimulus representations by specialized subsystems or activation processes (e.g., see Cowan, 1988, 1995, 1999). However, although these models assume that some sensory details are retained for longer periods of time, they do not explain the apparent contradiction between the classical findings of limitations in accessing sensory information and longer-term retention of these data. Our approach to this issue was based on everyday experience. One can recognize concrete objects (not just object categories) by sensory details even long after the object was last encountered. For example, we are often able to recognize our friends' voices over the telephone or shouting from another room, even though there are no known categorical properties distinguishing one voice from another. A model of the recognition situation may reveal important information about what type of sensory information is retained in the brain as well as about the circumstances that help in forming such memory representations. The auditory memory reactivation paradigm was designed to test these questions. By ba-

sing the test of "recognition" on an event-related brain potential (ERP) that can be measured independently of the subject's task, we avoided confounding factors stemming from the task and strategy of the subject, thus providing an unbiased measure of what sensory information is stored in the brain for longer time periods.

1.2 The Mismatch Negativity Event–Related Potential

Following a brief introduction to the ERP component used to test the recognition of sounds, the reactivation paradigm will be described in detail. We will show that with this paradigm, we are able to tap auditory sensory information that is resistant to decay and interference and discuss what processes and memory structures may underlie the reactivation phenomenon. Evidence will be provided showing that reactivated sound information can also be accessed in active discrimination tasks. In the discussion, we turn to the question of what kind of sound information is retained for longer periods of time. Finally, we relate our findings to two current models of working memory.

The ERP component involved in testing memory reactivation has been termed the mismatch negativity (MMN). MMN is elicited whenever a sound violates some regular aspect of the preceding sound sequence (for recent reviews, see Näätänen & Winkler, 1999; Picton, Alain, Otten, & Ritter, 2000). The elicitation of MMN requires the presence of some representation of the violated acoustic regularity. The auditory sensory information encoded in these regularity representations corresponds to that appearing in perception. MMN is elicited whether or not the subject's task is related to the test sounds. In fact, in the majority of MMN studies, subjects were engaged in some activity involving visual stimuli (e.g., they read a book, watched a movie, or performed some reaction task with visual stimuli) and were instructed to disregard the sounds presented to them (the "passive" condition). Thus MMN can be used to obtain a task-independent index of the retention of auditory information in the brain.

The MMN component typically peaks between 100 and 200ms from the onset of the regularity violation with negative polarity over the fronto-central scalp and positive over scalp locations above the opposite side of the Sylvian fissure. This is because the main cortical generators of MMN lie within or in the vicinity of the supratemporal plane with additional contribution from sources located in the frontal cortex (Halgren et al., 1995; Opitz, Rinne, Mecklinger, von Cramon, & Schrö-

ger, 2002). The simplest and most commonly used paradigm for obtaining MMN is the auditory oddball sequence. When a repeating sound (termed the "standard" sound) is occasionally exchanged for a different sound ("deviant" sound) MMN is elicited. However, MMN is also elicited by violations of far more complex auditory regularities (for reviews, see Näätänen et al., 2001; Winkler, 2003). The MMN wave can be delineated from other concurrent ERP components by subtracting from the ERP response elicited by the deviant stimulus the ERP elicited by some control sound. For a good assessment of the MMN response, the control sound should share as many features as possible with the deviant sound, but it should not itself elicit an MMN (i.e., it should be a stimulus that conforms to the regularities of the sound sequence in which it appears).

The current explanation of MMN elicitation suggests that incoming sounds are compared with extrapolations (sensory inferences) calculated from the representation of regularities detected in the preceding sound sequence. Sounds that mismatch these extrapolations activate the MMN-generating process (Winkler, Karmos, & Näätänen, 1996b). It is important to note that the presence of a sensory memory record of a sound in the brain is not a sufficient prerequisite of MMN elicitation; MMN is only elicited once some auditory regularity has been detected and a subsequent sound violates this regularity (Cowan, Winkler, Teder, & Näätänen, 1993; Sussman, Sheridan, Kreuzer, & Winkler, 2003a; Winkler, Schröger, & Cowan, 2001). Thus, stimulus change per se (e.g., two different sounds presented successively at the beginning of a sound sequence or within an ever-changing sequence of sounds) does not result in MMN elicitation (Cowan et al., 1993; Horváth, Czigler, Sussman, & Winkler, 2001; Winkler, 1996).

The MMN-generating process appears to be independent of top-down control (Rinne, Antila, & Winkler, 2001; Sussman, Winkler, & Wang, 2003b). It has been shown that in most cases, the MMN results obtained in the "passive" situation match those that can be obtained in the same paradigm with attention directed away from the sounds in a controlled manner (Sussman et al., 2003b; Winkler et al., 2003). However, it should be noted that some of the processes underlying the detection of auditory regularities can be modulated by top-down control and the outcome of these processes is also reflected in the elicitation and/or amplitude of the MMN response (Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Sussman et al., 2003b). The possible functions of the MMN-generating process are to initiate further processing of the deviant sounds (Näätänen, 1990) and/or to update the regularity representations that did not
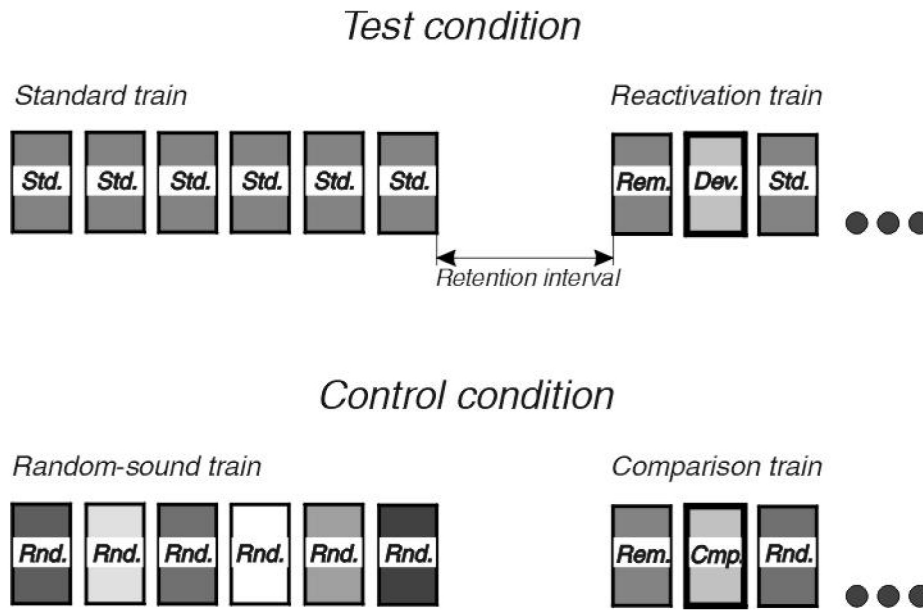
Test condition

Figure 1. Schematic illustration of an MMN reactivation paradigm. <u>Top panel</u>: The test condition. Standard trains consist of sounds or sound patterns, which conform to some regularity. Elements of these trains are termed "standards", marked with "Std" on the figure. Standard trains are followed by the retention interval and the reactivation train. The first element of the reactivation train is termed the "reminder" (marked "Rem"). It conforms to the regularities of the standard train (i.e., it is also a "standard"). Reactivation is tested by the second element of the reactivation train, which violates some regularity of the standard train. Thus it is termed "deviant", marked as "Dev". The reactivation train ends with a further standard or standards (their role is to provide a homogeneous context for the deviant). <u>Bottom panel</u>: One possible control condition. Standard trains are exchanged for trains in which the sounds or sound patterns vary in the regular feature of the standard trains (e.g., if the regularity used in the test condition is the constancy of tone frequency, then frequency is randomly varied in the control condition). The trains substituted for the standard trains are termed random-sound trains and their elements marked with "Rnd". Reactivation trains are exchanged for comparison trains that start with the same sounds as the reactivation trains (the reminder and a deviant, termed comparison tone and marked as "Cmp"). The rest of the train is made up of random sounds, again varying in the critical regularity of the test condition.

correctly predict the deviant (Winkler & Czigler, 1998; Winkler et al., 1996b).

2. Auditory Memory Reactivation

2.1. The Reactivation Paradigm

For testing reactivation with the MMN measure, at least two trains of sounds are needed (Figure 1, top panel). The first train sets up a regularity (tone repetition in Figure 1). It is termed the "standard train." The standard train is followed by the retention interval. The second train ("reactivation train") starts with a sound (or sound pattern) that conforms to the regularities of the standard train. This sound is termed the

"reminder." The second sound of the reactivation train then violates the regularity of the standard train. This sound is termed the deviant or position-2 deviant in paradigms testing MMN elicitation also in later positions of the trains. In some paradigms, each train served both functions: the first two sounds of the train tested reactivation with respect to the preceding train, whereas later sounds in the train set up the standard for the next train. If the (position-2) deviant sound elicits the MMN component, one can conclude that the regularity of the standard train was represented in the brain and this representation was available to the MMN-generating process. This is because, as was mentioned in the previous section, stimulus change alone does not activate the MMN-generating process. MMN can only be elicited if the deviant sound violates some detected regularity.[1] Figure 2 (top left

---

[1] The requirement of establishing a regularity in the standard train provides a good possibility for delineating the MMN component from other overlapping ERP responses. Figure 1 (bottom panel) shows the optimal control sequence. Standard trains are exchanged for trains that do not show the same regularity (termed "random-sound trains"). For example, if the deviant is set up to violate the common frequency of the standard-train tones, then a train of tones randomly varying in frequency can be used in the control condition. The "reminder" and the "deviant" are unchanged (compared with the test condition) but are again followed by
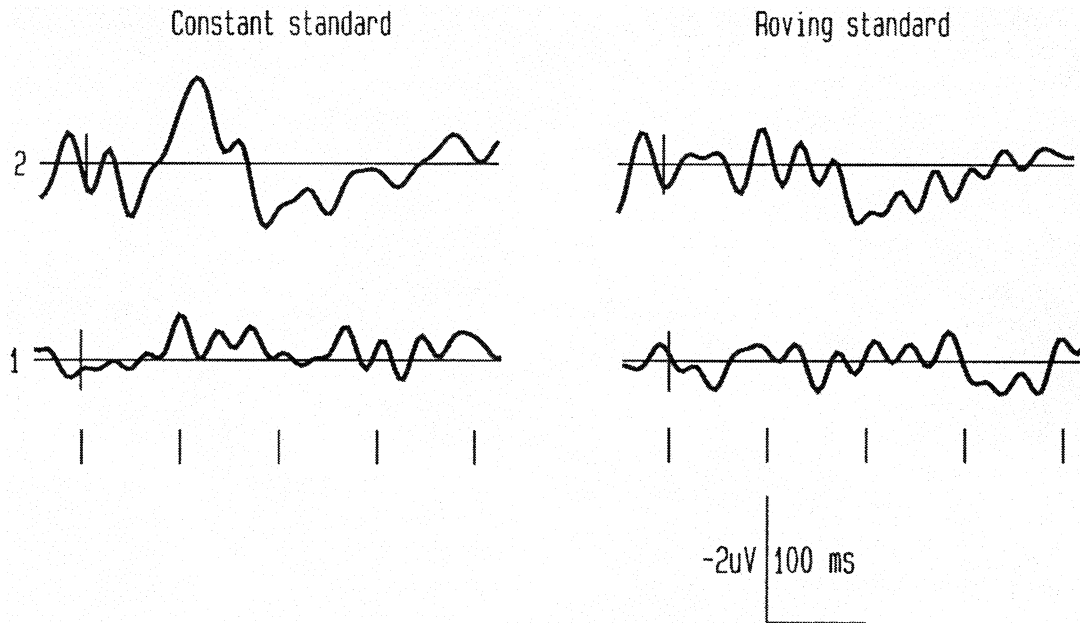
Figure 2. ERP difference waves (deviant minus same-position standard-tone responses) obtained by Cowan et al. (1993) at the frontal (Fz) electrode location in position 2 (reactivation, first row) and in position 1 (deactivation test, second row) of the tone trains. In the constant standard condition (left column), standard and deviant tones were fixed within the stimulus blocks. In the roving standard condition (right column), the frequency of both standard and deviant tones changed from train to train. Tone onset is at the crossing of the *x* and *y* axes. Calibration is marked at the lower right corner. Tick-marks at the bottom are spaced 100ms apart. The MMN response, which appears only in position 2 of the constant standard condition (upper left corner), is a negative wave peaking between 100and 200ms post-stimulus.

panel; adapted from Cowan et al., 1993) shows the MMN response elicited by a position-2 frequency-deviant tone.

However, the elicitation of MMN by a position-2 deviant, in and of itself, does not prove that reactivation occurred. One should also check whether the representation of the regularity of the standard train was still available to the MMN-generating process following the retention interval. That is, it should be tested whether a deviant presented in the first position of a stimulus train elicits the MMN. Figure 2 (bottom panels) shows that in Cowan et al.'s (1993) study no MMN was elicited by position-1 deviant tones. It has been repeatedly found that when trains are separated by a silent interval of 11 s or more, no MMN is elicited by deviant sounds presented at the beginning of a train (Cowan et al., 1993; Gaeta, Friedman, Ritter, & Cheng, 2001; Winkler et al., 2001). Therefore, when the retention interval is longer than 11 s, MMN elicited by position-2 deviants tells that the representation of the regularity of the standard train, which was not available to the MMN-generating process at the beginning of the reactivation train be-

came available by the time the second sound of the reactivation train was processed. Cowan et al. (1993) termed this phenomenon "memory reactivation" as they regarded it to be similar to the reactivation phenomena described by Rovee-Collier and Hayne (1987). Note, however, that the findings of Rovee-Collier and Hayne refer to learned actions rather than sensory information and that the timescale of their effect is much longer (several days) than that ever tested with the current reactivation paradigm.

The reminder is a critical element of the reactivation paradigm. Cowan et al. (1993) found that when the first tone of a train substantially differed from the tone repeated in the preceding train, it did not set up a subsequent "deviant" for MMN elicitation. In their "roving-standard" condition, Cowan et al. presented short trains of tones in which all but possibly one tone had the same (standard) frequency. This standard frequency, however, changed from train to train. In ca. 17% of the trains, a tone whose frequency differed from the current as well as from the previous standard frequency (deviant) appeared in the second position

random sounds. This setup ensures that the control-condition deviant will not elicit MMN while the rest of its ERP components match those elicited in the test condition. The first reactivation experiments used for comparison the response elicited by a standard sound that was presented in the same position as the deviant. More recent studies used the control described above.

of the train. The first tone of the train had the frequency of the new standard, which was different from that of the previous train. Position-2 deviants did not elicit MMN in this situation (Figure 2, top right panel). One could still argue that the roving-standard condition of Cowan et al. did not set up a stable regularity, because the standard frequency changed from train to train. However, Ritter and his colleagues (Ritter, Sussman, Molholm, & Foxe, 2002) have shown that reactivation occurs even if the standard frequency changes from train to train, when the reminder matches the standard of the preceding train. Winkler et al. (2002) narrowed the frequency range within which the reminder is effective (i.e., reactivation occurs). These authors found no MMN elicited by deviants following a reminder that differed only by 3% from the standard frequency. Winkler et al.'s result suggests that the representation against which the reminder is checked encodes features of the standard sounds with a resolution characteristic of sensory memory traces. Thus reactivation can indicate the existence of finely resolved auditory information in the human brain.

2.2. Characteristics of the Memory Involved in Reactivation

The first question is whether the characteristics the memory traces involved in reactivation are the same as those describing the classical notion of auditory sensory memory. Auditory sensory memory is subject both to decay (estimates ranging from 10 to 20 s) and to interference by similar sounds (Cowan, 1984). The first reactivation study (Cowan et al., 1993) tested retention intervals just beyond the most commonly accepted value for the duration of auditory sensory memory (11–15 s). It is thus possible that a weak residue of the auditory sensory memory trace of the repeated standard tone could have been reinforced by the reminder. However, in a more recent study (Winkler et al., 2002), the retention interval was set to 30 s. The standard was again a repeating tone and the deviant tone differed from it in frequency. All subjects showed reactivation in this situation, even though only one of them performed above chance level in discriminating the standard and deviant tones when the tones were separated by a 30-s silent interval.

Interference from similar sounds has also been tested (Winkler, Cowan, Csépe, Czigler, & Näätänen, 1996). Trains were composed of 12 tones. At least 5 of the first 6 tones were identical, whereas tones 7 through 12 (termed intervening tones) varied randomly in frequency. Deviants appearing within the first 6 tones also differed from the standard tones in frequency. Trains were separated by 5.9 s of silence. No MMN was elicited by the random-frequency intervening tones presented in the last 3 positions of the train (from position 10–12, i.e., the 4th–6th intervening tones). Because these tones deviated from the standard in the same way as the deviant did, the fact that no MMN was elicited by them at the end of the trains demonstrated that the frequency-regularity representation was no longer available to the MMN-generating process after the 1st through 3rd intervening tones, even before the onset of the silent interval that separated the trains. Therefore, if the sensory information involved in reactivation was vulnerable to interference from similar sounds, no MMN should be expected to be elicited by position-2 deviants (i.e., the deviants testing reactivation following the silent interval separating the trains). However, MMN was elicited by position-2 deviants in this situation. This result suggests that the memory involved in reactivation is not subject to the type of sensory interference that characterizes short-term auditory sensory memory.

The experiments reviewed above set up tone repetition as the standard. In most natural situations, however, sound sequences include substantial amount of variance and the regularities have to be extracted from the ever-changing input. One possible way to model such variability is to set random changes in some sound features while fixing the level of others. Ritter, Gomes, Cowan, Sussman, and Vaughan (1998) showed that constant tone-intensity was reactivated when tone-frequency was varied throughout the stimulus trains (including the frequency of the reminder tone) and vice versa, constant tone-frequency was reactivated when tone-intensity was varied. These results suggest that the memory representation involved in reactivation encodes constancies extracted from the variable input.

One could, however, argue that feature constancies may be detected directly by neurons sensitive to narrow ranges of auditory features (see, e.g., Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995). Furthermore, the type of sensory information whose recognition we intended to model with the reactivation paradigm is based on secondary, rather than primary sound features. For example, the timbre of a human voice that allows one to recognize a friend by his/her voice alone is characterized by the ratio between energy emitted in certain frequency ranges, rather than by the absolute pitch of the voice. A simple model of a secondary auditory feature has been constructed by Saarinen, Paavilainen, Schröger, Tervaniemi, and Näätänen (1992), who presented tone pairs, 90% of which were ascending in frequency (i.e., the frequency of the se-

cond tone of the pair was higher than that of the first tone). The absolute frequencies of the tones varied randomly throughout the sequences. Infrequent descending-pitched tone pairs and tone repetitions elicited the MMN response (for a study controlling all aspects of pitch ascension, see Paavilainen, Jaramillo, Näätänen, & Winkler, 1999).

Korzyukov, Winkler, Gumenyuk, Alho, and Näätänen (2003) investigated whether this pitch-ascension regularity can be reactivated similarly to regularities based on primary sounds features. Seven tone-pairs of ascending pitch varying randomly in their absolute frequency levels were presented in the standard trains. The retention interval was followed by a train starting with one ascending-pitch (reminder) and a descending-pitch (deviant) tone pair. (Control sequences were composed according to the scheme shown on the bottom panel of Figure 1: pitch direction of the tone-pairs varied randomly in the random-sound trains.) A significant difference was observed between the responses elicited by the deviant and the corresponding comparison tone-pair in the MMN latency range. The scalp distribution of this potential difference matched that of the MMN component elicited by descending-pitch tone pairs presented infrequently amongst frequent ascending-pitch tone pairs. This result demonstrated that regularities based on secondary sound features can be reactivated. Thus we can conclude that the reactivation paradigm can indeed model real-life situations in which recognition occurs on the basis of acoustic subtleties.

2.3. Interpretation of the Reactivation Phenomenon

Cowan et al. (1993) offered two alternative explanations of their findings of "memory reactivation." One hypothesis suggests that during the retention interval, the sensory memory traces involved in detecting acoustic deviance enter a dormant state (i.e., a state in which they cannot be directly accessed). The reminder activates the corresponding dormant memory trace(s), bringing them back to immediate memory (and thus allowing the detection of the following deviants). The alternative explanation assumes that, although the memory traces required for deviance detection are present and accessible, they are not consulted by the MMN-generating process because the retention interval causes a context change. That is, the sounds presented after the relatively long retention interval are not initially considered to be a continuation of the preceding sound sequence, but rather the start of a new sound group, whose regularities are yet to be de-termined. In this explanation, the function of the reminder is to provide a link to the previous train, reinstating its context as current and the related regularities as relevant to the processing of the new sounds.

The lack of MMN elicitation by the first sound of the trains following the retention interval should thus be taken either as a sign of the dormancy of the memory traces or as a sign that a change of context have taken place. Based on the results of a behavioral test and the analysis of the MMN studies known at the time, Ritter and his colleagues (2002) argued that, in just 10–15 seconds, auditory sensory memory traces do not decay beyond usefulness. Evidence supporting Ritter et al.'s conclusion has been obtained by Winkler and his colleagues (2001; see also Gaeta et al., 2001), who found that deviant tones may not elicit MMN at the beginning of a short train when the trains are separated by a silent interval of just 7 s duration. The trains consisted of 4 tones of uniform stimulus duration (equiprobably 100 or 300 ms) and were delivered equiprobably with a stimulus onset asynchrony (SOA; onset-to-onset interval) of 0.5 or 7 s. The silent interval separating successive trains was always 7 s. The ERP response to the first tone of those trains in which tone duration differed from the preceding train was compared with that elicited when tone duration matched with the preceding train. The change in tone duration elicited MMN in ca. half of the subjects when the within-train SOA of the preceding train was 0.5 s. In contrast, the same deviants following the same silent interval (the between-train interval was always 7 s) elicited MMN in all subjects when the within-train SOA of the preceding train was 7 s (i.e., equal to the inter-train interval). This result cannot be explained on the basis of sensory memory alone, because, at the time when the deviant tone was delivered, the sensory memory trace of the standard tone must have been stronger when the preceding train were delivered with the short SOA than when it was delivered with the long SOA. This is because more standard tones were delivered with short than the long SOA within the last ca. 10 s preceding the deviant tone. Therefore, we must assume than the sensory memory trace of the standard tone was present and accessible in both conditions at the time when the deviant tone was delivered. Thus, this result demonstrates that the lack of MMN cannot be taken as proving that no sensory memory trace can be accessed by the MMN-generating process. Winkler et al. (and also Gaeta et al.) interpreted their result in terms of the context change hypothesis. On the basis of these results, Ritter et al. (2002) favored the context reactivation (or reinstatement) explanation of the MMN reactivation phenomenon.

However, since then, Winkler et al. (2002) showed reactivation after a much longer (30 s) retention interval and that reactivation occurred also in those subjects who could not discriminate the standard and deviant tones when these were separated by 30 s of silence. The latter result is all the more important since Keller et al. (1995) found a small effect of tone rehearsal on tone comparisons performance for over-10-s retention intervals. This may explain why Ritter et al. (2002) as well as our own test (see the next section) found significant residual memory after 11–15 s in active tone comparison tasks. However, the finding of reactivation in subjects who could not perform the comparable discrimination task suggests that reactivation may involve both of the processes brought up by Cowan et al. (1993): The reminder connects the new sounds to a previous context (reinstatement) as well as reactivating the possibly dormant memory representations that describe the regular characteristics of this context.

2.4. Reactivation in an Active Tone–Comparison Experiment

The comparisons between results obtained with MMN and in behavioral studies (discussed in the previous section) make it imperative to test reactivation in an active paradigm that is analogous to the ones tested in the passive situation. Although the reactivation paradigm was intended to model situations in which recognition occurs without voluntary effort, one would assume that reactivation should also occur in situations in which the subject actively tries to maintain some sensory information. This was tested in two experiments in which subjects were required to judge whether a test tone was the same or different compared with a previously presented standard tone. The two experiments differed only in whether the standard tone was presented only once (1-Standard Experiment; 21 participants, 9 male, 16–23 years of age, 19.4 years mean age) or six times in a row with 0.75 s SOA (6-Standard Experiment; 25 participants, 13 male, 18–31 years of age, 22.4 years mean age).

2.4.1. Stimuli and Procedure

Eighteen sets of tones and a burst of white noise, bandpass filtered between 300 and 1500 Hz, were generated. The duration of all sounds was 250ms (including 5ms rise and 5ms fall times), their intensity 70 dB (SPL). Each tone set consisted of 3 tones separated in fre-

quency by proportionally equal steps of 2.5, 3, 3.5, 4, or 5% (see later). (For later reference, the middle tone of the sets will be denoted as "A" as it was always the first to be presented; the lower-pitched one will be denoted as "B," the higher as "C," and the aperiodic noise as "N"). The frequency of the middle (A) tone of the lowest set was 400 Hz; the middle tones of neighboring sets were separated by 6% in frequency. Each trial presented tones from only one set. Tone sets appeared in the trials in a randomized order and with equal probability. Trials started with the standard tone(s), which was always the middle tone of a tone set ("A"), followed by a silent retention interval, which was, with equal probability, 11, 12, 13, 14, or 15 s. The sound following the retention interval was selected with equal (25%) probability from the three tones of the set and the noise-burst. These second tones were considered to be reminders (A valid; B, C, and N invalid). Following the reminder by an SOA of 0.75 s, the test tone was delivered. The test tone was equiprobably chosen from the three tones (A, B, and C) of the set selected for the trial. Thus, 12 types of trials were presented (the order of the sounds being standard, reminder, test): AAA, AAB, AAC, ABA, ABB, ABC, ACA, ACB, ACC, ANA, ANB, and ANC. Each trial type occurred 18 times, each time based on a different tone set (216 trials, overall). Subjects were instructed to press a button if they thought that the test tone was identical to the standard tone(s) and a different button if they thought that the two tones were different. The reminder (the sound preceding the test tone) was introduced to them as a warning signal preparing them for the delivery of the test tone. They were informed that the warning sound carried no information about the task and so they should not rely on it in their judgment. The instructions emphasized the requirement of correct responses and placed no time pressure on giving the response. Subjects were also motivated by a scheme of performance-dependent bonus payments to do the task as best they could. Subjects started each trial when they felt ready for it.

Prior to starting the main test, we established the frequency difference at which subjects could reliably (> 90%) discriminate two tones of slightly different frequencies. Subjects were presented with pairs of tones separated by a 2-s SOA. The two tones of the pair were either identical or slightly different from each other in frequency. In separate stimulus blocks (10 pairs per block), the frequency difference was either 2.5, 3.0, 3.5, 4.0, or 5.0%. Absolute tone frequencies varied across the 18 preselected frequency sets (see above). Testing started with the highest (5%) frequency separation. After each successful stimulus

block (a block in which at least 9 of the 10 responses were correct), frequency separation was decreased. Frequency separation was increased after two successive blocks in which performance did not reach the criterion level. Testing was finished when a frequency separation was found at which the subject performed at the desired level, but had to turn back twice from the one-step lower frequency-separation level. On average, subjects of the 1-standard experiment needed 4.1 %, whereas subjects of the 6-standards experiment needed 3.8 % frequency separation for reliable discrimination performance. Frequency separation between the B and A and A and C tones was set up in the main experiment to equal the level established in the preliminary testing. Subjects then received training in the task of the main experiment. Two blocks of 18 trials each were presented to them with the retention interval set to 3.5 s and feedback given after each response. All other parameters were identical to the corresponding ones in the main experiment.

2.4.2. Model for Analyzing the Results

Performance was analyzed with the help of a mathematical model[2] of performance that assumed additivity between the following effects:

(1) Memory ("M") – Performance based on the sensory memory trace present at the time the warning (reminder) tone was presented.

(2) Reactivation ("R") – Only present on AAX type of trials (X can be either A, B, or C); increases performance.

(3) Interference from the warning tone ("I") – Decreases performance by degrading the residual memory trace. Interference was assumed to be zero for noise (ANX) trials. This assumption may not be accurate. As a consequence, "M" will be slightly underestimated.

(4) Strategy to answer according to the relationship between the warning and the test tone ("S") – Using this strategy boosts performance on some types of trials, such as ABC, but degrades performance on others, such as ABB. The frequency separation between the warning and the test tones was larger in the ABC and ACB trials than in any of the other trials. This may have increased the likelihood in these types of trials that answers were given on the basis of the relationship of the warning and the test tones. By assuming that "S" was equal in all trial types, we may thus underestimate "S" for the ABC and ACB type of trials and, as a consequence, underestimate the size of the reactivation ("R") effect.

(5) Bias to answer "equal" over answering "different" ("B") – Bias boosts performance on some types of trials, such as ABA, but degrades performance on others, such as ABB. This tendency was assumed to depend on the relationship between the standard tone and the warning sound. Therefore three separate "B" values were considered, one for the AAX trials, another for the ABX and ACX trials, and the third for the ANX trials: Bias variables were named Ba, Bbc, and Bn, respectively.

The number of variables (effects) in the model equaled the number of independent measurements. Therefore, the effect values (the amount by which each effect contributed to performance) could be unambiguously determined, separately for the 1- and 6-Standard experiments.

2.4.3. Results and Discussion

Figure 3 presents the results of both active reactivation experiments (hit percentages according to trial categories on the top panel and the calculated model values on the bottom panel). Although the memory factor was responsible for the largest segment of performance, reactivation proved to be significant in both experiments ($p < 0.05$ in the 1-Standard Experiment and $p < 0.01$ in the 6-Standard Experiment; one-group Student's t tests). The strategy and interference factors were found to be significant only in the 6-Standard Experiment ($p < 0.01$ and $p < 0.05$, respectively). No significant differences were found between the two experiments for any of the model factors (dif-

---

[2] The formal model of performance is as follows. P(AXY) stands for the performance in the AXY trials (X can be A, B, C, or N and Y can be A, B, or C); for other abbreviations, see the text. Equations are numbered from (1) to (7).
(1) $P(ANA) = M + Bn$
(2) $[P(ANB) + P(ANC)]/2 = M - Bn$
(3) $P(AAA) = M + R + S + Ba - I$
(4) $[P(AAB) + P(AAC)]/2 = M + R + S - Ba - I$
(5) $[P(ABA) + P(ACA)]/2 = M - S + Bbc - I$
(6) $[P(ABB) + P(ACC)]/2 = M - S - Bbc - I$
(7) $[P(ABC) + P(ACB)]/2 = M + S - Bbc - I$

Tone Comparison

Model Values
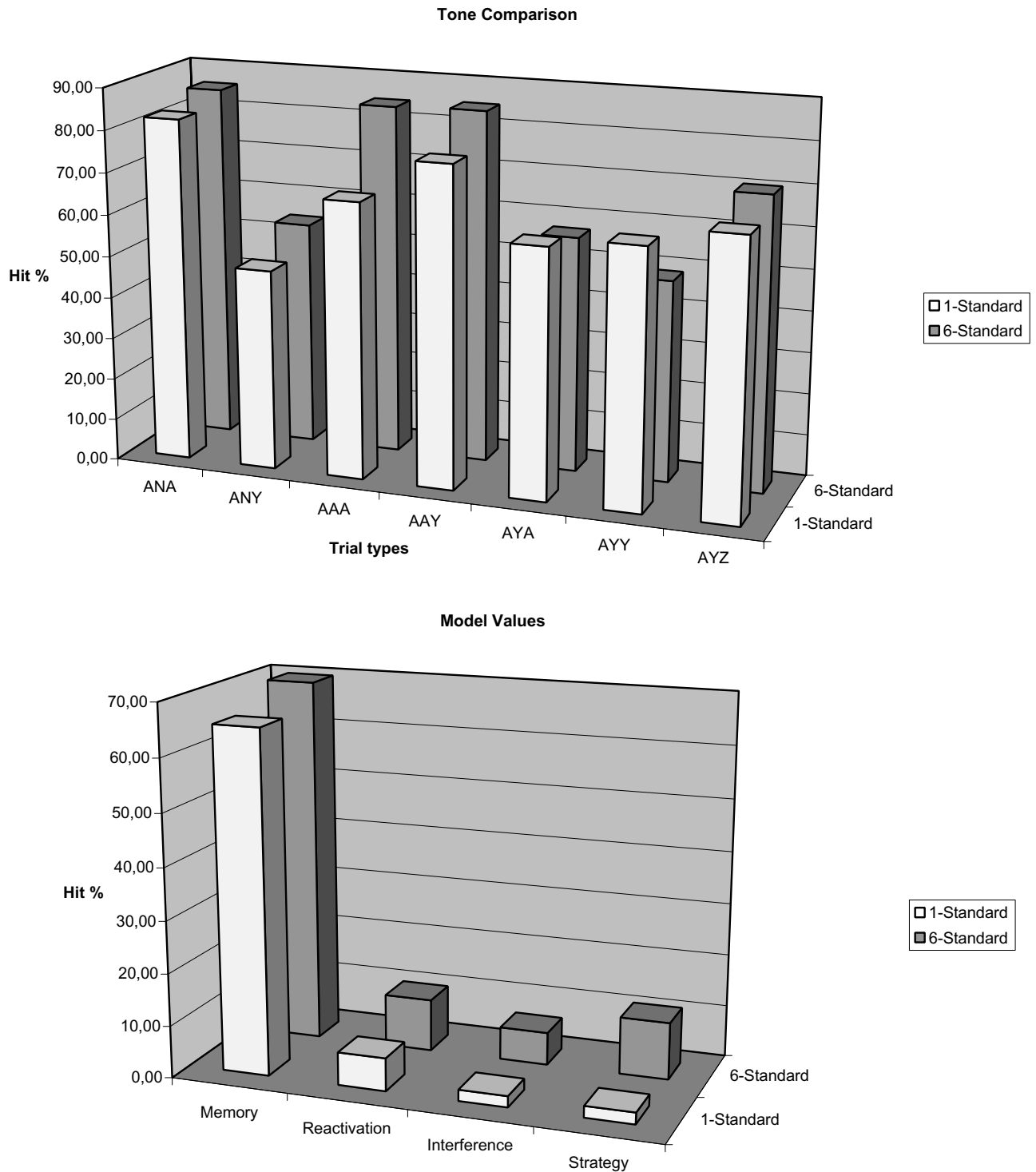
Figure 3. Active reactivation experiments. <u>Top panel</u>: The percentage of correct responses (hits), separately for the 1-Standard (front row) and 6-Standard (back row) experiments and sorted by trial categories. The 12 trial types (see the text) were analyzed in 7 categories created according to the different effects included in the model. ANY presents the average performance in the ANB and ANC trials, AAY the AAB and AAC trials, AYA the ABA and ACA trials, AYY the ABB and ACC trials, and AYZ the ABC and ACB trials. <u>Bottom panel</u>: Contribution of the modeled effects (in hit percentage) to the performance in the two experiments. Note that the interference factor decreases performance (all other factors, when their value is positive, increase performance).

ference in the strategy factor was marginally significant: $F(1, 44) = 3.80$, $p = 0.0575$; one-way between-group ANOVA). In the 6-Standard but not in the 1-Standard Experiment, reactivation was positively correlated with memory (Spearman rank correlation 0.44, $p < 0.05$), whereas strategy was negatively correlated with memory (-0.43, $p < 0.05$). These results suggest that reactivation occurs also when subjects actively maintain auditory sensory information, although its effect is relatively small as long as the residual memory is still sufficiently strong (i.e., in the present situation, over 60% of the performance can be attributed to residual memory, whereas only ca. 10% to reactivation).

3. General Discussion: A Reevaluation of Sensory Memory

3.1. What Kind of Memory Representations Are Reactivated?

MMN elicitation is based on a memory representation describing auditory regularities, not just auditory sensory memory traces. This has been shown by the results of those studies that tested violations of various nonrepetitive regularities (for a review, see Näätänen et al., 2001). However, these memory representations are not independent of the concrete sound feature levels even in the most "abstract" cases. For example, Paavilainen, Simola, Jaramillo, Näätänen, and Winkler (2001) presented subjects with tones varying in frequency and intensity. Most tones conformed to a rule, which was, in separate stimulus blocks, either "the higher the frequency the higher the intensity" or "the higher the frequency the lower the intensity". Infrequent tones violating the standard rule (high-frequency soft and low-frequency loud tones or high-frequency loud and low-frequency soft tones, depending on the rule) elicited MMN. This result suggests that the memory representation involved in MMN generation encoded the abstract feature-conjunction regularity. One aspect of the results, however, suggested that the representation of the regularity was not fully independent of the actual levels of the relevant auditory features. The MMN amplitude was somewhat higher in response to deviants that fell farther from the center of the feature distribution of the regular sounds. (However, no regular tone elicited the MMN, not even the ones with extreme levels in both features.) The MMN-amplitude differences suggest that the standard tones of medium frequency and medium intensity were regarded as a prototype within the abstract regularity.

The picture emerging from the MMN literature is that the representations of the stimulation must include not only regularities, but also certain feature values that serve as reference points or anchors. An anchor point would allow an abstract rule or regularity to be translated into specific feature values. That way, the same regularity could be associated with different anchors in different stimulus situations, allowing for the efficient storage of information and the possibility to maintain alternative descriptions of the same sequence of sounds. Simultaneous representation of multiple redundant regularities describing the same sound sequence has been demonstrated for MMN generation (Horváth et al., 2001). Horváth and his colleagues presented a regular sequence of tones alternating in pitch (ABABAB..., where A and B are two tones differing only in pitch). The presence of representations for different redundant regularities was tested by presenting deviants that violated one possible description (rule) while conforming to a different description of pitch alternation. It was found that memory representations of at least one "local" rule (A tone is always followed by B and vice verse) and one "global" rule (every second tone is A and every other is B) are simultaneously maintained and incoming sounds are checked against them in parallel. Further results suggested that possibly also representations of more general versions of alternation are kept active at the same time (rules, such as Higher tone – Lower tone – Higher tone – Lower tone ... and alternation with variable interstimulus intervals). If, as was shown by Horváth et al. (2001), the auditory system maintains multiple redundant representations even for simple acoustic regularities, concise abstract regularity descriptions actualized with a minimal amount of sensory data offer an economical form of information storage.

The hypothesized structure of regularity representations suggests that we can reactivate sensory information that was encountered within a context with distinctive regular characteristics. A stimulus that conforms these regularities serves as a reminder, activating the regularity representations and thus bringing the corresponding context to immediate memory. Reactivation of the characteristic features of an object could serve as the basis of its recognition. Features of the reactivation process are in perfect correspondence with everyday experience of object recognition. We can easily recognize objects appearing in their usual context. The more experience we have with the given object and the more distinctive its sensory features the more likely that we can recognize it.

The characteristics of the memory representations as shown by the reviewed reactivation experiments,

i.e., resistance to decay and interference, encoding of higher-order regularities extracted from acoustical variance, are compatible with the assumed characteristics of long-term memory records. Thus it seems possible that in the auditory modality, the regularity representations indexed by the MMN component form the link between immediate sensory memory and long-term storage of sensory information.

3.2. What Distinguishes Auditory Sensory Memory from Categorical Memory?

The MMN and new behavioral results discussed above help to establish several points about memory for acoustic stimuli. First, the regularities of the acoustic pattern, including grouping and organization of acoustic information, are important for memory. Second, aspects of these regularities can be stored in long-term memory and reactivated later. Given these results, several issues arise with respect to behavioral research on memory and the models that have been based on it. First, does it still make sense to think of this information as especially "sensory," as opposed to categorical, in nature? Second, what types of overall models of processing can represent all of the results?

In the following we shall re-examine the evidence supporting the distinction of sensory memory traces form categorical memory storage for each of the four critical qualities mentioned in the introduction: (1) sensory memory does not depend on attention, (2) it is modality-specific, (3) it includes information that is finer than any meaningful set of categories, and (4) it is short-lived. How do these qualities apply to the retention of acoustic regularities?

3.2.1. Attention–Independence

A traditional objection to the attention-independence of sensory processing comes from findings showing attentional modulation of neural signals associated with early afferent processing. Results showing that mid-latency ERP components, such as the auditory P50 wave are attenuated when attention is directed away from the stimuli have been interpreted in terms of "sensory gating" (e.g., Guterman, Josiassen, & Bashore, 1992). It has been suggested that the afferent flow of sensory information is under control from the prefrontal cortex, a structure linked with the voluntary direction of behavior (Knight, Staines, Swick, & Chao, 1999). The notion of sensory gating may be compatible with the attentional spotlight theory. Sup-

porting results for the auditory modality have been obtained in experiments investigating the detection of auditory targets as a function of the target's distance from a precued (expected) location. A graded decrease of accuracy and slower reaction times have been found with increased distance between the target and the precued location (Arbogast & Kidd, 2000; Mondor & Zatorre, 1995; Rorden & Driver, 2001). Comparable amplitude gradients have been found both for some of the early, exogenous (N1) as well as for later, endogenous (MMN, P3) ERP responses (Arnott & Alain, 2002; Teder-Salejärvi, Hillyard, Roder, & Neville, 1999).

Effects in sound organization provide another line of argument against the attention-independence of auditory sensory memory. Bregman (1990) suggested that several heuristic processes analyze the auditory input in parallel. These processes provide alternative solutions to breaking down the complex acoustic input into coherent sequences of sound. When two strong alternatives of sound organization emerge from the initial analysis (ambiguous auditory scenes), one can voluntarily choose between them. Since the auditory information stored about sounds depend on how they are organized (Dowling, 1973), ambiguous auditory scenes allow voluntary modulation of auditory sensory memory. In a similar vein, Sussman et al. (2002) found that informing subjects about the large-scale structure of a sound sequence determined what sounds are detected as deviants in a sound sequence. Sussman et al. presented a tone sequence composed of a repeating tone pattern (AAAABAAAAB ...; where B was higher in pitch than A). Tones were presented at an intermediate (700ms) SOA, because previous studies showed that at shorter SOAs (100ms), the sequence is unambiguously represented as a repeating tone pattern (Sussman, Ritter, & Vaughan, 1998), whereas at longer SOAs the large-scale regularity is not automatically detected (Scherg, Vajsar, & Picton, 1989). Of the A tones, 2.5% were exchanged for a lower-pitched (C) tone. When subjects were instructed to press a response button for the rare low-pitched C tones MMN was elicited by the B tones, which were relatively rare (appeared in 20% of the time) compared with the A tones. In contrast, when subjects were informed about the regularly repeating tone pattern and were instructed to press the response button whenever the pattern was broken (that is, to the same C tones), the B tones did not elicit MMN. The lack of MMN by the B tones indicates that the tone sequence was represented in terms of the repeating tone pattern. In this case, the B tone is part of the repeating pattern and, therefore, it does not violate a regularity. Thus it appears that, at least in certain cases, auditory memory

can be influenced by top-down processes. Moreover, some authors suggest that even the unambiguous cases of sound organization require attention (e.g., Carlyon, Cusack, Foxton, & Robertson, 2001), although this issue is still controversial (see, e.g., Winkler et al., 2003).

Taken together, this literature does call into question whether complex acoustic patterns are held in a way that is independent of attention; but it also calls into question the more fundamental assumption that sensory memory is held in a manner that is independent of attention. That never has been rigorously examined, and the aforementioned study by Keller et al. (1995) did find that attention had an effect on memory for tone frequencies necessary for comparisons within a musical category.

3.2.2. Modality-Specificity

Regarding the second criterion of sensory memory, its modality-specificity, here the results seem unequivocal. Frankish (1985) and Cowan, Saults, Elliott, and Moreno (2002) compared memory for grouped and ungrouped lists of spoken or printed verbal items. The result in both studies was that grouping has a much larger, beneficial effect in the auditory modality. In fact, Cowan et al. found that, with visual presentations, grouping could be accomplished just as well on trials in which the stimuli themselves were presented at a steady pace as on trials in which the stimuli were temporally grouped into 3 clearly distinct sets of 3 digits. Both presentation schedules produced serial recall results for the 9-digit lists that were slightly scalloped in a manner, suggesting that the 9 digits had been mentally grouped accordingly. (The stimulus grouping on some trials apparently induced a similar mental grouping on the remaining trials with visual presentation.) In contrast, with auditory presentation of digits, the outcome for grouped lists was far superior to the outcome for lists of digits presented at a steady pace. It is as if the acoustic modality, unlike vision, inherently carries timing information that cannot be ignored within the mental representation. In accordance with this notion, Näätänen & Winkler (1999) suggested that the time-line of acoustic stimulation serves as the core for integrating the outcome of the various auditory feature analyzers into a unitary representation of the auditory stimulus: "We suggest that the critical step between the stage of fragmentary stimulus information (maintained in the feature traces) and the emergence of the auditory stimulus representation is the synthesis of the static stimulus features with the temporal envelope of the stimulus event. An auditory stimulus cannot be fully described by static features alone. Therefore, in the formation of an auditory stimulus representation, the extracted features must be aligned with the passage of time." (Näätänen & Winkler, 1999, p. 848).

3.2.3. Resolution

The third traditional hallmark of sensory information is that it is more fine-grained than categorical information. The question is whether this fine-grained information is present within the same memory representation that stores extended acoustic patterns. Evidence to that effect is that the ability to detect a difference between two tones differing less than a musical note category is affected by the pattern of prior tone presentations. Cowan, Saults, and Nugent (1997) varied not only the time between tones to be compared, but also the time between trials. It was found that when trials were further apart, performance improved. One way to explain that result is that, when trials are too close together, the first tone in the current pair may be grouped in memory with tones from the preceding trial. So, it appears that there is no such thing as a simple, pure acoustic memory for a single tone separate from the pattern of tones that has formed.

3.2.4. Duration

The fourth typical characteristic of sensory memory was its short life. However, we already have discussed evidence that there is memory for acoustic properties that lasts a long time, such as information about a friend's voice. There is abundant evidence for long-term auditory memory that allows a recall advantage over visual presentation, even after an intervening distracting task, provided that the items to be recalled also are separated by other items. Research on long-term modality effects recently was reviewed by Gardiner and Cowan (2003). Also, Cowan, Saults, and Nugent (2001) reanalyzed the evidence on tone comparisons by Cowan et al. (1997) and found that, under certain circumstances in which the separation between trials was the largest, it was unclear if there was any loss of information at all as a function of the time between tones to be compared. In other words, the "decay" of auditory sensory information across several seconds may actually be a matter of the tone becoming more and more confusable with tones from previous trials, as opposed to an actual loss of sensory persistence as has typically been assumed. However, this dramatic conclusion awaits further testing.

Although these findings suggest that sensory information lasts in memory rather than decaying rapidly, there are some apparent contradictory findings. Some studies have shown that the final serial position of acoustically presented lists carries information that is especially vivid and may be represented differently from items presented in any other position of the list. There has been some debate on this point (cf. Balota & Engle, 1981; Bloom & Watkins, 1999). However, Cowan, Nugent, Elliott, and Saults (2000) found a clear distinction of the final serial position in a developmental study of memory for digit lists that were ignored during their presentation while a silent game involving rhyming pictures was played. Occasionally, the rhyming game was interrupted and the computer keyboard was to be used to recall the most recent spoken list. The delay between the end of this last list and the recall cue was 1, 5, or 10 s. For most serial positions, children of two ages and adults all forgot the items at the same rate across the variable retention interval. For the final serial position, though, the result was strikingly different than for any other serial position. Young children (in second grade) forgot this final item much more rapidly than subjects in the older groups.

There is another finding suggesting that there may be a special form or function of memory that is specific to the final item (although see Bloom & Watkins, 1999 for evidence that it is not quite that specific). It has been obtained in studies of the suffix effect, in which a final list item that is not to be recalled (e.g., the word "go") interferes with memory for items at the end of the list (Crowder & Morton, 1969; Morton, Crowder, & Prussin, 1971). Balota and Ducek (1986) presented a suffix immediately after a spoken list or after a 20-second delay and found that there was a suffix effect in either case, but that the similarity between the voice of the speaker of the list and suffix mattered only if the suffix was presented immediately. This again suggests that there is some specific acoustic information that is very fragile and some broader pattern of information that is more durable. It is unclear if the specific acoustic information disappears as a function of absolute time or, more in keeping with other evidence we have presented, disappears as a function of the shifting context; the specific acoustic information apparently would lose its contextual relevance especially quickly over time.

3.3. The "Regularity–Record Plus Anchor" Hypothesis

The theoretical need, then, is to tie together findings suggesting that acoustic information is retained for a long time and the findings suggesting that the final item may be of special importance. Information on long-term retention includes, for example, the mismatch negativity results showing that patterns are taken into account and that rapid decay does not describe performance well (e.g., Winkler et al., 2001), and behavioral results on long-term modality effects (e.g., Gardiner & Cowan, 2003) and sensory memory stability (Cowan et al., 2001). Information on the special importance of the most recent item includes the role of this item as an anchor, or as a reminding or reactivating stimulus, in mismatch negativity procedures (Cowan et al., 1993; Korzyukov et al., 2003; Winkler et al., 1996) and, behaviorally, the developmental difference in memory decay for the list-final item only (Cowan et al., 2000) and the tendency of modality and suffix effects to be largest at the end of the list.

One way to combine these factors theoretically is with the notion that acoustic memory is a record of regularities, and that the last item serves as an anchor for this record. As an analogy, every member of an orchestra has a mental record of the music that is to be played, but still needs a common pitch to anchor the memory and tune the instrument before the musical piece is played. To carry the analogy further, the melody can be held in long-term memory but the anchoring note must be played on the spot for it to be of any use (except for musicians with the absolute pitch ability). If an intruder came into the room during tuning and played a deviant note, the process would be corrupted and the correct anchor probably would have to be repeated. That is analogous to the use of a reactivating reminder in the mismatch negativity procedure and to the corrupting effect of a suffix in list-recall procedures.

Characteristics of the suffix effect are compatible with the above-suggested structure of auditory memory (i.e., regularity plus anchor). Most studies of the suffix effect show that it can occur even when the suffix remains the same from trial to trial. However, the situation is different when list items are separated by distracting tasks, making them temporally very distinct and producing long-term modality effects. Under that situation, a suffix effect eliminating the auditory modality advantage occurs, but only if a different suffix is used on every trial (Glenberg, 1984). In the short term, the most recently presented acoustic item may have a special vividness that is susceptible to interference from any other sound. In the long term, though, the most recent item may carry acoustic information that reminds the subject of the rest of the acoustically presented list. When that item was a suffix that was not to be recalled, it may detract from the

power of the list-final item to serve as a unique anchor for the list.

3.4. Considering some Current Memory Models

Last, we can ask the question of what established models can or cannot handle these results. The foremost model of temporary memory, currently, is the working-memory model of Baddeley (1986). That model postulates the existence of two buffer stores, a phonological buffer and a visuospatial buffer. Recently, a short-term episodic buffer also has been postulated (Baddeley, 2000). However, in our view, none of these buffers make sufficient accommodation for specifically sensory information. The phonological buffer cannot be used to explain the auditory modality superiority effect. To explain it, it seems necessary to postulate the existence of sensory memory stores in addition to the more abstract, phonological, and spatial representations that are highlighted in Baddeley's model.

In the similar but alternative model of Cowan (1988, 1995, 1999), sensory memory and categorical memory both are considered to reflect activated subsets of the long-term memory system (though new links are formed between items that are attended concurrently, and these new links then become part of the long-term memory system). Results showing fast task-independent effects of information stored in long-term memory records on auditory change detection are fully compatible with Cowan's model. For example, phonetic category boundaries and prototypes have been shown to affect the MMN response elicited in ignored sequences of speech stimuli (Aaltonen, Eerola, Hellstrom, Usipaikka, & Lang, 1997; Näätänen et al., 1997; Winkler et al., 1999). In Cowan's model, the modality-specific effects are handled by assuming that there will be interference between the activated representation of a recently presented item and a new stimulus when the two share similar features. Cowan's (1988, 1995, 1999) model does not specifically explain why the auditory modality superiority effect occurs; nor does it specify the nature of memory for regularities in the auditory memory system. Yet, it seems more open to the possibility of separate acoustic and phonological information sources than does Baddeley's (1986, 2000) model.

Given the way in which the expectations of the earlier models of processing have failed, it is perhaps not surprising to learn that some researchers have promoted models in which there is no distinction at all between shorter-term and longer-term memory (e.g.,

Nairne, 2002). Such models are well suited to accommodate the finding that memory for acoustic regularities is long lasting, and to accommodate similarities between the results of ostensibly short-term and long-term memory phenomena. However, such models do not yet appear to have a ready explanation for differences that are observed between shorter-term and longer-term memory phenomena. One issue here is what we mean by the term "reactivation." It carries with it the assumption that there is such a thing as activation. If there is no distinct form of memory in the short term, then there is no such thing as temporary memory activation. Instead, what we term "activated" would actually be "relevant to the present context." In that case, reactivation effects would have to be more accurately portrayed as reminder effects.

Ultimately, new modeling efforts will be needed to explain precisely how it is that regularities can be stored in acoustic memory, and to determine whether some minimal amount of attention is needed to assist in the formation of the memory for regularities that is stored and later can be reactivated by a reminder stimulus.

References

Aaltonen, O., Eerola, O., Hellstrom, Å., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America, 101,* 1090–1106.

Arbogast, T. L., & Kidd, G. (2000). Evidence for spatial tuning in informational masking using the probe-signal method. *Journal of the Acoustical Society of America, 108,* 1803–1810.

Arnott, S. R., & Alain, C. (2002). Stepping out of the spotlight: MMN attenuation as a function of distance from the attended location. *NeuroReport, 13,* 2209–2212.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory.* Oxford Psychology Series #11. Oxford: Clarendon Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Balota, D. A., & Duchek, J. M. (1986). Voice-specific information and the 20-second delayed suffix effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 509–516.

Balota, D. A., & Engle, R. W. (1981). Structural and strategic factors in the stimulus suffix effect. *Journal of Verbal Learning and Verbal Behavior, 20,* 346–357.

Bloom, L. C., & Watkins, M. J. (1999). Two-component theory of the suffix effect: Contrary findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1452–1475.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press.

Broadbent, D. E. (1958). *Perception and communication.* New York: Pergamon Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. J. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 115–127.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191.

Cowan, N. (1995). *Attention and memory. An integrated framework.* Oxford: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, UK: Cambridge University Press.

Cowan, N., Nugent, L. D., Elliott, E. M., & Saults, J. S. (2000). Persistence of memory for ignored lists of digits: Areas of developmental constancy and change. *Journal of Experimental Child Psychology, 76,* 151–172.

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46,* 153–177.

Cowan, N., Saults, J. S., & Nugent, L. D. (1997). The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone pitch comparisons. *Psychonomic Bulletin & Review, 4,* 393–397.

Cowan, N., Saults, S., & Nugent, L. (2001). The ravages of absolute and relative amounts of time on memory. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. Surprenant (eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 315–330). Washington, DC: American Psychological Association.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Short- and long-term prerequisites of the mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 909–921.

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology, 26,* 274–284.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 472–478.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage. *Perception & Psychophysics, 5,* 365–373.

Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (eds.), *Short-term memory* (pp. 107–151). New York: Academic Press.

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics, 14,* 37–40.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46,* 372–383.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language, 24,* 200–209.

Fujisaki, H., & Kawashima, T. (1971). A model of the mechanisms for speech perception - quantitative analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute of the Faculty of Engineering, University of Tokyo, 30,* 59–68.

Gaeta, H., Friedman, D., Ritter, W., & Cheng, J. (2001). The effect of perceptual grouping on the mismatch negativity. *Psychophysiology, 38,* 316–324.

Gardiner, J. M., & Cowan, N. (2003). Modality effects. In J. H. Byrne, H. Eichenbaum, H. Roediger III, & R. F. Thompson (eds.), *Learning and Memory* (2nd edition) (pp. 397–400). New York: Macmillan.

Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 16–31.

Guterman, Y., Josiassen, R. C., & Bashore, T. R. Jr (1992). Attentional influence on the P50 component of the auditory event-related brain potential. *International Journal of Psychophysiology, 12,* 197–209.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liegeois, C., Chauvel, P., & Musolino, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalography & Clinical Neurophysiology, 94,* 191–220.

Horváth, J., Czigler, I., Sussman, E., & Winkler, I. (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences. *Cognitive Brain Research, 12,* 131–144.

Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 635–645.

Knight, R. T., Staines, W. R., Swick, D., & Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychologica, 101,* 159–178.

Korzyukov, O., Winkler, I., Gumenyuk, V., Alho, K., & Näätänen, R. (2003). Processing abstract auditory features in the human auditory cortex. *Neuroimage, 20,* 2245–2258.

Massaro, D. W. (1970). Perceptual processes and forgetting in memory tasks. *Psychological Review, 85,* 411–417.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 387–409.

Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology, 91,* 169–190.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13,* 201–288.

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

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

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125,* 826–859.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage, 15,* 167–174.

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

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology, 38,* 359–365.

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

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260.

Rinne, T., Antila, S., & Winkler, I. (2001). Mismatch negativity is unaffected by top-down predictive information. *NeuroReport, 12,* 2209–2213.

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

Ritter, W., Gomes, H., Cowan, N., Sussman, E., & Vaughan, H. G. Jr (1998). Reactivation of a dormant representation of an auditory stimulus feature. *Journal of Cognitive Neuroscience, 10,* 605–614.

Ritter, W., Sussman, E., Molholm, S., & Foxe, J. J. (2002). Memory reactivation or reinstatement and the mismatch negativity. *Psychophysiology, 39,* 158–165.

Rorden, C., & Driver, J. (2001). Spatial deployment of attention within and across hemifields in an auditory task. *Experimental Brain Research, 137,* 487–496.

Rovee-Collier, C., & Hayne, H. (1987). Reactivation of infant memory: Implications for cognitive development. *Advances in Child Development & Behavior, 20,* 185–238.

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.

Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience, 1,* 336–355.

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

Sussman, E., Sheridan, K., Kreuzer, J., & Winkler, I. (2003a). Representation of the standard: stimulus context effects on the process generating the mismatch negativity component of event-related brain potentials. *Psychophysiology, 40,* 465–471.

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

Sussman, E., Winkler, I., & Wang, W. J. (2003b). MMN and attention: Competition for deviance detection. *Psychophysiology, 40,* 430–435.

Teder-Salejärvi, W. A., Hillyard, S. A., Roder, B., & Neville, H. J. (1999). Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials. *Cognitive Brain Research, 8,* 213–227.

Winkler, I. (1996). Necessary and sufficient conditions for the elicitation of the mismatch negativity. In C. Ogura, Y. Koga, & M. Shimokochi (Eds.) *Recent Advances in Event-related Brain Potential Research* (pp. 36–44). Amsterdam: Elsevier.

Winkler, I. (2003). Change detection in complex auditory environment: Beyond the oddball paradigm. In J. Polich (Ed.), *Detection of Change: Event-Related Potential and fMRI Findings* (pp. 61–81). Boston: Kluwer Academic Publishers.

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

Winkler, I., & Czigler, I. (1998). Mismatch negativity: deviance detection or the maintenance of the "standard.". *NeuroReport, 9,* 3809–3813.

Winkler, I., Karmos, G., & Näätänen, R. (1996b). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research, 742,* 239–252.

Winkler, I., Korzykov, O., Gumenyuk, V., Cowan, N., Linkenkaer-Hansen, K., Alho, K., Ilmoniemi, R. J., & Näätänen, R. (2002). Temporary and longer retention of acoustic information. *Psychophysiology, 39,* 530–534.

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

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

Winkler, I., Sussman, E., Tervaniemi, M., Ritter, W., Horváth, J., & Näätänen, R. (2003). Pre-attentive auditory context effects. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 57–77.

Adress for correspondence

István Winkler
Institute for Psychology
Hungarian Academy of Sciences
P.O. Box 398
H-1394 Budapest
Hungary
Tel. +36 1 354 2296
Fax +36 1 354 2416
E-mail winkler@cogpsyphy.hu