An EEG-based Approach for Evaluating Audio Notifications under Ambient Sounds

Yi-Chieh Lee[1], **Wen-Chieh Lin**[1], **Jung-Tai King**[2], **Li-Wei Ko**[2,3], **Yu-Ting Huang**[1], **Fu-Yin Cherng**[1]
[1]Dept. of Computer Science, [2]Brain Research Center, [3]Dept. of Biological Science and Technology
National Chiao Tung University, Taiwan
be341341@gmail.com, wclin@cs.nctu.edu.tw, {jtchin2, lwko.brc, tinahuang79, fufu22710}@gmail.com

ABSTRACT

Audio notifications are an important means of prompting users of electronic products. Although useful in most environments, audio notifications are ineffective in certain situations, especially against particular auditory backgrounds or when the user is distracted. Several studies have used behavioral performance to evaluate audio notifications, but these studies failed to achieve consistent results due to factors including user subjectivity and environmental differences; thus, a new method and more objective indicators are necessary. In this study, we propose an approach based on electroencephalography (EEG) to evaluate audio notifications by measuring users' auditory perceptual responses (mismatch negativity) and attention shifting (P3a). We demonstrate our approach by applying it to the usability testing of audio notifications in realistic scenarios, such as users performing a major task amid ambient noises. Our results open a new perspective for evaluating the design of the audio notifications.

Author Keywords

audio notifications; usability testing; human cognition; electroencephalography (EEG); mismatch negativity

ACM Classification Keywords

H.1.2 User/Machine Systems: J.3Life and Medical Sciences

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

Auditory icons, earcons, and warning sounds have been widely used in computers, consumer electronics, and traffic-control devices. Audio notifications also play critical roles in assistive technologies for the visually impaired; they are a powerful tool for human computer interfaces and can be used to assist users to work more effectively [3]. Generally speaking, the design goals of audio notifications are that they should be short and clear, and allow users to recognize and understand their meanings easily [3, 29]. However, a notification's effectiveness at attracting users' attention is also important. An inappropriate audio design may distract users' concentration, even to the point of danger. Hence, a good audio notification design should take into account not only users' preferences but also the notification's perceptual effects upon users' cognitive states.

Existing approaches for evaluating the preference and effectiveness of audio notifications often rely on questionnaires, or measures of users' performance, such as response time and hit rate [26]. While these approaches provide means of evaluating various audio notifications, they are limited with regard to analysis of users' cognitive states, such as alertness level and attention shifting. Previous studies [14, 24] have shown that humans' behavioral performance is a rather indirect indicator of their cognitive state and may not be consistent with it at all. Hence, failure to assess cognitive state in usability testing might lead to an incorrect evaluation [24]. Despite the wide applications of audio notifications, there has been little research aimed at developing objective measurement of perceptual or cognitive activities for audio-notification evaluation. Moreover, as an audio notification often plays a supporting role in our interaction with machines, its influence to the major task on auditory perception and cognitive state should not be ignored.

Recently, brain-sensing technologies have been used to evaluate users' cognitive state directly [7, 28] and research attempts have been made to integrate them into usability testing in the HCI field [12, 24]. Nevertheless, previous work of using EEG or other brain-sensing technologies [12, 24, 31] focused on investigating users' workload or visual tasks, and these techniques have rarely been used to measure auditory tasks. As auditory stimulation is attention-grabbing [3] and humans can hardly avoid hearing it, it is thus very important to assess the perceptual effects of sound and users' cognitive state when we evaluate audio notifications for HCI design.

In this paper, we propose to apply EEG technology to evaluate human auditory perception and attention allocation in usability testing. We exploit two types of neural responses, called mismatch negativity (MMN) and P3a, which can be detected in our brains when we hear novel sounds. Depending on the characteristics of a sound and a user's cognitive state, different degrees of the MMN and P3a responses would be induced. While various studies in cognitive science have

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

http://dx.doi.org/10.1145/2556288.2557076

investigated MMN and P3a, most still use simplified sounds, such as pure tones and beeps, in their experiments. Although some research used repetitive tones and natural sounds [20], there is still a significant gap between MMN/P3a studies and HCI applications.

Therefore, we conducted two experiments to determine whether MMN and P3a could be new indicators to enhance the usability testing of audio notifications. If so, this would allow audio designers to use our approach to objectively assess the impact of their audio design upon users, and to improve their design accordingly. In addition, we suggest some possible future applications of MMN and P3a research to HCI.

This paper makes three primary research contributions:

(1) *Extending prior studies and verifying that MMN and P3a can be induced by complex audio stimuli.* This extension is critical for applying MMN and P3a to HCI research as MMN and P3a experiments have rarely been conducted on stimuli in daily life such as audio notifications and ambient noises.

(2) *Suggesting a new method for objectively evaluating audio notifications.* Existing paradigms for verifying the effectiveness of audio notifications mostly require subjects to pay attention to the stimuli. However, Sussman et al. [30] have shown that if subjects expect audio notifications to occur, the impact on their sound perception is the same but the influence on their attention state is clearly different. Therefore, we introduce an objective approach for audio evaluation that can be conducted without users' attention, i.e., a passive and unobtrusive way to evaluate audio notifications.

(3) *Demonstrating that MMN and P3a could enhance the audio usability testing in scenarios where the evaluated task is supporting*. Users operate across a wide variety of environments in their daily lives, and this certainly influences the perceptual effects of audio notifications. Moreover, users undertake tasks that require different amounts of mental workload, and this too would tend to affect the evaluation results. We introduce an appropriate approach for considering these complex factors in usability testing of audio notifications.

RELATED WORK AND BACKGROUND

Audio Notifications

Auditory icons and earcons are audio notifications commonly used in consumer products and by mobile services. Gaver [13] first introduced auditory icons, which map a specific event to a sound heard in everyday life such that users can easily remember the icon without learning. In contrast to auditory icons, earcons [2] are nonverbal messages represented by abstract musical tones that users have to learn and memorize.

Brewster [3] suggests that auditory interfaces should avoid annoying sounds because they may disturb other people in the same environment. Furthermore, one should avoid basing the meaning of an audio notification on its intensity, since a loud sound may grab too much of a users' attention and could even cause a delay in their handling the problem. Hence, it is necessary to carefully evaluate the perception and effectiveness of audio designs.

Graham [11] compared the effectiveness of auditory icons and earcons as vehicle collision warnings. He found that auditory icons had shorter reaction time but higher risk of false-positive responses. Edworthy [5] also suggested that the real-world sounds are more suitable for monitoring tasks, while earcons are more appropriate for the warning systems as they can effectively attract the user's attention. Lemmens et al. [17] studied whether audio notifications help subjects react faster in a picture categorization task. Their results showed that both auditory icons and earcons reduce subjects' response time, as compared to the quiet condition.

Conventionally, response time, hit rate, and questionnaires have been used to evaluate the effectiveness and impact of auditory icons and earcons. However, these prior audio-evaluation approaches have some limitations. Fagerlönn et al. [8] indicated that non-expert users do not necessarily feel comfortable about giving suggestions or feedback on a detailed level. Other studies [24, 28] also pointed out that performance measurement could not take into account users' changing cognitive state, such as mental workload during an interactive process. In fact, previous studies [1, 18] had suggested that audio-notification designers should consider the context of use, especially when the users are under complex and high mental-workload conditions. Nevertheless, measuring human perception is difficult because users' experiences and perceptions vary from time to time [18]. To help designers objectively measure the perceptual effectiveness of their audio designs, we need a method to complement prior evaluation approaches.

Brain-Sensing for Evaluation

Several studies [28, 31] have applied brain-computer interface (BCI) to HCI. Zander and Kothe [33] proposed *passive* BCI, which records human cognitive state without the purpose of voluntary control. Using *passive* BCI can enrich HCI and help monitor actual user states during interaction. To understand the perceptual influence and the degree of attention shifting associated with particular audio notifications, we need a more efficient evaluation method. One possible solution is to make use of physiological indicators, such as functional near-infrared spectroscopy (fNIRS) or event-related potentials (ERPs) via EEG device. fNIRS is a non-invasive imaging method involving the quantification of chromosphere concentration resolved from the measurement of near-infrared light attenuation, or temporal or phasic changes. Although fNIRS is an adequate tool for monitoring cognitive states, it is not suitable for quick perceptual tasks [24]. In contrast, as ERPs are brain responses directly related to specific sensory, motor, or cognitive events, they provide higher temporal resolution in analyzing the various stages of neural processing associated with related behaviors [20]. Therefore, in this study, we utilize ERPs to analzye the effectiveness of various audio notifications.

Mismatch Negativity

An ERP component well known to be related to sound processing is the MMN, which is a change-specific negative potential over fronto-central sites. The latency of MMN is about

100-250 ms, and it is elicited by a subject's preattentive discrimination between a recurring standard sound and a deviant sound [20]. For example, when subjects are seeing a movie and hearing an oddball sequence of sounds, in which a rare deviant sound is interspersed among a series of frequent standard sounds, the repetitive standard sound forms a memory trace and MMN will be evoked if a new (deviant) sound does not match this auditory memory. This means that humans' auditory system can use preattention to distinguish the differences between two sounds, including intensities, pitches, locations, and durations [23, 16]. Moreover, MMN reflects not only the physical characteristics of deviant sounds (e.g., intensity, frequency, duration), but also more abstract representations of complex auditory rules, such as melody, speech, and rhythm [20].

The amplitude and latency of MMN depend on the differences between the deviant and standard sounds, with larger differences inducing larger amplitudes and shorter latencies [20]. MMN amplitude may also positively correlate with hit rate [22], where the correlation between MMN response and behavioral performance is investigated across various frequencies of the same sound. These studies suggest that MMN is a good indicator for audio perception and an objective measurement for the discriminability of a sound.

P3a
A fronto-central positive component, P3a, often occurs around 200 ms to 300 ms after the stimulus onset, i.e., somewhat later than MMN, and is related to the orienting responses towards deviant stimuli [6]. When deviant stimuli attract more of a subject's attention, the amplitude of P3a becomes larger. Although increasing the difference between the deviant and standard stimuli often enlarges the MMN and P3a amplitudes at the same time [20], this tendency holds under proper conditions [22]. Previous research has indicated that P3a would also be modulated according to different cognitive states such as workload and attention [25, 30]. Based on these studies, MMN is more likely related to involuntary attention switching, while P3a is more likely to reflect the level of active attention orienting.

Oddball Paradigm
A common task used in MMN and P3a studies is the oddball paradigm, in which a series of auditory or visual stimuli with different occurrence probabilities are used to assess neural reactions to unpredictable events [20]. An auditory oddball paradigm usually includes a frequent tone as standard stimuli and a less frequent tone as deviant or novel stimuli. When infrequent stimuli are inserted in the series of standard stimuli, MMN and P3a – which as we have seen are associated with novelty detection and attention-switching towards deviant events – will be evoked [6], and the occurrence probability of frequent and infrequent stimuli will influence the level of MMN and P3a. Typically, the cumulative duration ratio of deviant stimuli is controlled at 10%-30% of the cumulative duration of standard stimuli [20]. However, some studies still found stable MMN when 50% of deviant stimuli were used [27].

RESEARCH GOALS
Our goal is to explore the possibility of applying EEG to the evaluation of the effectiveness of audio notifications under more realistic conditions that are suitable for HCI applications. In the first experiment, we used ambient noises and audio notifications as stimuli to gain an understanding of how these auditory stimuli affect users' perception. Though prior studies have found that MMN can be evoked under complex standard stimuli such as repetitive tones or natural sounds [20], these cannot represent real conditions. Thus, a novelty of our experiment design was that we used ambient noises instead of repetitive sounds as the standard stimuli, and audio notifications as the deviant stimuli. This brought our experimental conditions closer to real-world scenarios, thus potentially extending the applications of MMN and P3a for audio evaluation.

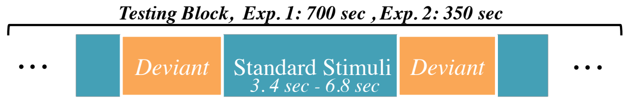
In the second experiment, we explored the benefits of MMN and P3a for analyzing subjects' perception of sounds and their attention shifting while evaluating audio notifications in a more complex and realistic situation, where subjects simultaneously performed multiple tasks with auditory and visual modalities. We manipulated each subject's cognitive state by controlling the difficulty of the game (the card game Pairs in this case), to test if human perception of sound and attention shifting could be affected by users' mental workload, and if these influences resulted in the deflection of MMN and P3a.

METHODS
We modified the oddball paradigm to evaluate the effectiveness of six audio notifications in two sets of ambient noises.

Deviant Stimuli: Audio Notifications
Six audio notifications were used as deviant stimuli: Dev1 (Do, 500 ms), Dev2 (Do, 500 ms), Dev3 (Tri-tone, 500 ms), Dev4 (Ding, 500 ms), Dev5 (Horn, 700 ms) and Dev6 (Bell, 700 ms). All deviants were presented via loudspeakers. The intensity of each deviant was set at 72 dB SPL, except that of Dev2 which was set at 78 dB SPL. Dev1 and Dev2 were the same digitally generated piano chords, but with different intensities, to allow us to assess the effect of intensity on audio notification. Dev3-Dev6 were four of the default audio notifications used on Apple iPhones to notifying their users about incoming messages.

Standard Stimuli: Ambient Noises
Two sets of ambient noises recorded at real-life venues were used as the standard stimuli: the first was recorded in a club, and included conversations of males and females, some music, and background noise; the other was recorded on a city street and included traffic sounds and other street noise. The intensities of the two ambient noises were controlled to between 70 to 75 dB SPL, which was a comfortable level for the subjects in our test scenarios. The experimental setting and information regarding auditory stimuli are presented in Figure 1.

Procedure
We set a period of 700 sec as a testing block (Figure 1). The total time of all deviants appearing in each block was controll-



	Standard Stimuli	Deviant Stimuli
Experiment 1	Club Ambient	Dev1 – Dev6
	City Street Ambient	Dev1 – Dev6
Experiment 2	Club Ambient	Dev1, Dev2
	City Street Ambient	Dev1, Dev2

Figure 1. List of experimental procedures and auditory stimuli. All deviant stimuli are randomly scattered in a testing block. The occurrence probability of each deviant is equal.

ed so as to not exceed 30%, i.e., 210 sec. The interval between each deviant stimulus was at least 3.4 sec. To avoid subjects generating expectations of the audio notifications, and to collect enough trials, the interval between any two consecutive deviant stimuli was randomly varied in the range of 3.4 to 6.8 sec. In each block, all six deviant stimuli were randomly scattered with equal probability of occurrence and played as an overlay to the ambient noise while the chosen set of standard stimuli (i.e., either city street noise or club noise) was played continuously. Figure 1 shows our paradigm design. We took the EEG data recorded between each pair of consecutive deviants as the standard trials and those recorded while each audio notification appeared as the deviant trials. We obtained approximately 120 deviant trials in each testing block. At the beginning of each block, standard stimuli (ambient noise) would be played for 10 sec, data from which was not recorded.

EEG Recording and Data Analysis

We used a non-invasive EEG cap as measurement equipment. The EEG was recorded using a NeuroScan system with 32 Ag-AgCl electrodes and a bandpass filter of 0.01-100 Hz. Although 32-channel EEG data were recorded, we only used single-channel (Fz) ERP information for analysis as previous studies had shown that MMN has its maximal amplitude at Fz which is usually treated as the representative location of the MMN source [20, 22]. Additionally, for practical considerations of usability testing, it would be advantageous if fewer EEG channels (implying a cheaper, portable EEG device) are proved to be sufficient for a quick evaluation. We adopted an open-source Matlab toolbox, EEGLAB, to process EEG data, which was digitized with a 1000 Hz sampling rate and bandpass filtered at 1-50 Hz off-line. The epochs used for averaging were 600 ms long, starting 100 ms before and ending 500 ms after the stimulus onset. The pre-stimulus period (100 ms) was set as a baseline. All epochs with voltage variations exceeding 100μV were automatically rejected; the rejection rate varied from 5%-10%. Then, all epochs of the same standard (or deviant) stimuli were averaged. For MMN and P3a analysis, the response to the standard stimuli was subtracted from the response to each deviant stimuli. The peak amplitudes and latency of MMN and P3a were measured separately from the most negative peak occurring at 100-220 ms and the most positive deflection within 220-320 ms at the location of electrode Fz (Figure 2). Mauchly's test and Greenhouse-Geisser corrections were applied to verify if the data violated the assumption of sphericity. Repeated-measures (RM) ANOVAs were conducted, and the Tukey's method for multiple comparisons was applied in post-hoc tests.

EXPERIMENT 1: AUDIO EVALUATION BY MMN AND P3A

The goal of this experiment was to extend prior MMN and P3a studies to the evaluation of audio notifications in the HCI field, by bringing the experiment design closer to real-life scenarios. Figure 1 indicates the procedure and auditory stimuli involved this experiment.

Participants and Device

Thirty healthy, right-handed subjects (12 females; 20-25 years old) with normal or corrected-to-normal vision participated in this experiment. None of the subjects had any history of brain disease, drug use, or hearing problems. None had any musical expertise. After completing the entire experiment (about 2 hours), subjects were rewarded with 16 USD for their time. We used an LCD monitor (22 inches, 1920×1080 pixels) to show videos and visual tasks. Audio stimuli were presented via loudspeakers (Altec Lansing 2.0 ch, VS2620), which were placed at a distance of 60 cm in front of the subjects, at 30° to their left- and right-hand sides. A decibel-meter was used to adjust the intensity of audio stimuli before all experiments.

Task

There were two sessions intended to test the relationship between the MMN, P3a and behavioral performance in experiment 1. All subjects attended both sessions.

Session 1. We only recorded the subjects' behavioral data for the audio notifications. The subjects were seated in a comfortable chair and instructed to visually stare on a white cross sign in the center of the screen before the experiment began. Once the experiment started, they watched a subtitled silent video pre-chosen by themselves, and continuously heard an ambient noise. They needed to respond via the keyboard as fast as possible once they heard any of the six audio notifications.

Session 2. The subjects watched the same silent video along with the same ambient noise as in Session 1, but they were asked to ignore all auditory stimuli while their EEG data were recorded during the entire session.

Each session comprised four testing blocks: two using club ambient noise and two using city street ambient noise as standard stimuli. Between any two consecutive blocks, subjects had three minutes to rest. The order of the four blocks for each subject were randomly arranged. During each session with a given subject, at least 40 trials were collected for each audio notification in a session.

Results of Experiment 1

Session 1: Behavioral Performance

Behavioral responses were defined as 'hit trials' if subjects pressed the keyboard within 2 sec after a deviant stimulus had played. We computed the average reaction time and hit rate for each audio notification. Only the reaction time of the hit

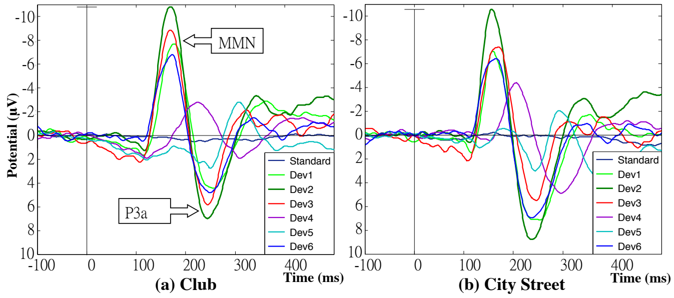


Figure 2. MMN (100-220 ms) and P3a (200-320 ms) curves in the context of two ambient noises: (a) Club; (b) City Street.

trials were used to compute the averaged reaction time. We performed a two-way RM-ANOVA (2 ambient noises × 6 deviant stimuli) to analyze the relationship between each audio notification. Mauchly's test indicated that the sphericity was not violated.

Figure 3(a)(b) shows the average reaction time (RT) associated with each audio notification within each type of ambient noise. The results did not reveal significant effects of the type of ambient noises (F=.056, p=.812). We found Dev5 had statistically significant differences with the other deviants under both types of ambient noises (p<.05). Additionally, Dev4 differed significantly from Dev2 (p=.02) and Dev3 (p=.01) in the context of city-street ambient noise. There were no statistically significant differences between any two of Dev1, Dev2, Dev3 and Dev6 in either ambient-noises context.

Figure 3(c)(d) shows the average hit rate of each audio notification within each type of ambient noise. There was a main effect of deviants on hit rate (F=4.32, p<.002). It can clearly be seen that Dev4 and Dev5 had lower hit rate than the other deviants, especially in the context of club ambient noise. There was a statistically significant difference between Dev4/Dev5 and the other four deviants (p<.05). The average hit rates of Dev1, Dev2, Dev3 and Dev6 in the club background were higher than their hit rates in the city-street background, while Dev4 and Dev5 had lower hit rates in the club background. This implies that subjects' behavioral performance on audio notifications might be influenced by ambient noises.

Session 2: MMN and P3a

An RM-ANOVA was used to determine if audio stimuli induced different MMN and P3a potentials. Mauchly's test indicated that the assumption of sphericity was not violated. We used 95% as the confidence level to determine the statistically significant difference.

Club Ambient Noise. Figure 2(a) shows the grand-peak ERP curves elicited by each deviant stimulus and the standard stimulus (club ambient noise) on the Fz channel. Each curve was obtained by averaging the ERPs of all subjects' trials. There was a main effect of different audio notifications on MMN amplitude (F=288.8, p<.0001). Except for Dev5, all deviants elicited a significantly larger negative deflection than the standard stimuli in the 100-220 ms interval after the stimuli onset. This showed typical MMN responses. As our results closely replicated those of previous studies [32], it seems appropriate to use MMN amplitude and latency to explore the discriminability of audio notifications in the context of different ambient noises.

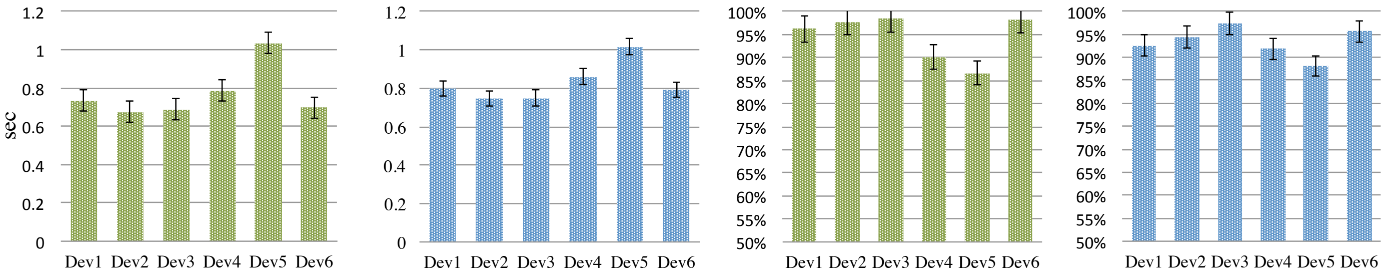
The MMN amplitude of each deviant is shown in Figure 4(a), which reveals a ranking of the different deviants' amplitudes: Dev2 > Dev3 > Dev1 > Dev6 > Dev4 > Dev5 (p<.01 in all post-hoc comparisons). We observed that Dev3 and Dev4 had the same intensity and duration, but their MMN amplitudes had a statistically significant difference (p<.001). A similar result was also found with regard to Dev5 and Dev6 (p<.001). These results revealed that two deviants' MMN could be very different despite their intensities and durations being the same. According to previous studies, following the occurrence of MMN, a positive ERP component P3a would be elicited over the frontal-central area (Fz channel) [25]. Our results also clearly replicated this P3a response around 220-320 ms, as indicated in Figure 2(a). Figure 4(c) plots the P3a amplitude of the six deviants. There was also a main effect of different deviants on P3a amplitude (F=36.12, p<.001). One can clearly distinguish their differences. Dev1, Dev2, Dev3 and Dev6 showed more positive deflection than Dev4 and Dev5. No statistically significant difference was found between Dev4 and Dev5 (p=.39), on the one hand, and on the other, Dev1 and Dev6 (p=.87).

City-Street Ambient Noise. The ERP curves of all deviants in the context of city-street ambient noise also exhibited typical MMN responses (F=127.46, p<.0001), with the exception of Dev5 (Figure 2(b)). Figure 4(b) shows the ranking of the deviants based on their MMN amplitudes: Dev2 > Dev3 ≥ Dev1 ≥ Dev6 > Dev4 > Dev5, where p<.05 in all post-hoc comparisons except Dev1/Dev3 (p=.14), and Dev1/Dev6 (p=.52). Additionally, there was a main effect of different deviants on P3a amplitude (F=93.6, p<.001). Dev2 had the largest positive potential, while Dev1 (p=.005) and Dev6 (p=.008) had greater positive potential than Dev3.

Discussion

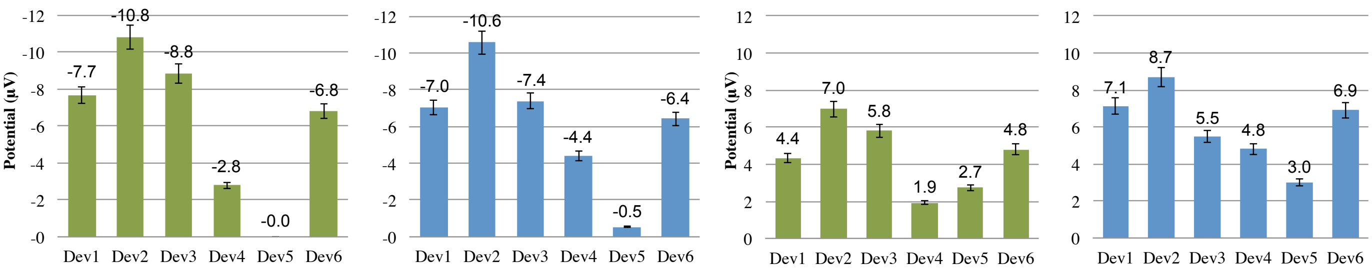
MMN and P3a Reflect Multiple Properties of a Sound. By manipulating only the intensity of two of the deviant stimuli – Dev1 and Dev 2 being the same sound, but at 72 dB and 78 dB, respectively – we found statistically significant differences between both their MMN and P3a amplitudes. When an audio notification was louder, the subjects were more likely to notice it and paid more attention to it. This finding replicated previous research [23, 20] that the physical characteristics of different sounds can be reflected by the ERP components, MMN and P3a.

Although physical characteristics of sounds, such as intensity and frequency, can greatly affect brain response, they are not the only factors influencing MMN. For example, Dev3 and Dev4 had almost the same intensity and duration, but their MMN amplitudes differed greatly. This supports the idea that MMN is not only influenced by the intensity and duration of a sound [20]. The same phenomenon was also observed in the case of different notifications that had the same decibel level or duration, e.g., Dev5 and Dev6.



(a) Reaction Time (Club) (b) Reaction Time (City Street) (c) Hit Rate (Club) (d) Hit Rate (City Street)

Figure 3. (a) and (b) are average reaction times associated with each audio notification, computed from the hit trials of all subjects, while (c) and (d) are the average hit rates of each audio notification.



(a) MMN Amplitude (Club) (b) MMN Amplitude (City Street) (c) P3a Amplitude (Club) (d) P3a Amplitude (City Street)

Figure 4. (a) and (b) are MMN amplitudes in the context of the two different sets of ambient noise, while (c) and (d) are P3a amplitudes in the two different sets of ambient noise.

Some audio notifications had an effect on the latency of MMN and P3a. Both Dev4 and Dev5 had longer latency when compared with the other four deviants, regardless of the type of ambient noises (Figure 2). This long latency suggest that the deviant required more time to be detected and perceived. This could be because of other properties of the sound, such as melody or timbre, and more experiments will be needed to make a determination. We also found a ranking order among the MMN amplitudes of the different audio notifications (Figure 4(a)(b)). This shows that audio notifications could be discriminated by brain response. However, the ranking order of the deviants by MMN amplitude was not exactly the same as their ranking by P3a. This indicated that MMN and P3a may not be mutually dependent. For instance, Dev1 and Dev3 had similar MMN potentials in the city-street background, but Dev1 had larger P3a potential than Dev3. This is not surprising, as MMN and P3a actually reflect auditory perception and involuntary attention shifting, respectively [20]. In other words, MMN shows how well a subject detects a sound while P3a shows how much the subject's attention is shifted towards it. Thus, it can be supposed from the above example of Dev1 and Dev3 that these two deviants were detected by brains at a similar level, but they led to different degree of attention resource allocation.

Behaviorial Performance v.s. MMN and P3a. In general, the behavioral data (Figure 3) can provide a simple evaluation of audio notifications, but they cannot discriminate clearly between all audio notifications well. On the other hand, MMN and P3a clearly revealed the ranking order of different audio notifications. This suggests that they can provide additional information for evaluating audio notifications. In particular, MMN and P3a can reflect the impact of audio notifications on subjects' perception and attention shifting. Given that this deeper analysis of auditory perception and attention is possible, a better evaluation approach to the evaluation of audio notifications is likewise possible.

In fact, MMN and P3a can help to analyze behavioral performance. For example, Dev5 had a longer reaction time and a lower hit rate than the other deviant stimuli in Figure 3. The reason for this cannot be learned from behavioral performance alone, but MMN and P3a may provide some clues. Indeed, Dev5 induced the lowest MMN potential in both ambient-noise backgrounds. This implies that it is difficult for users to discriminate Dev5 from ambient noises. A similar observation can also be made with regard to Dev4. These findings are in agreement with those of previous studies [22], with a relatively more simple experimental setting (i.e., modulating frequency of the same audio stimuli). They suggested that MMN amplitude might positively correlate with the hit rate and negatively correlate with the RT.

Effect of Ambient Noises. Our six audio notifications also exhibited different ERP responses with the two types of background sounds. We found that P3a amplitudes of four audio notifications (i.e., except Dev3 and Dev5) exhibited significant differences between the two ambient noises, while the MMN amplitudes of Dev3, Dev4, and Dev5 also were significantly different depending on ambient noises. This implies that the subjects' perception of a deviant sound and attention shifting to the deviant sound, were both influenced by the type of background sounds. Furthermore, though audio notifications can be discriminated amid different ambient noises, their discriminability might be influenced by the discrepancy between them and those ambient noises. In fact, previous studies have shown that ambient noises would influence not only the performance of human [1, 18] but also creative cognition [19]. Our findings indicated that EEG method is a useful tool for verifying audio perception in the context of different types of ambient noises.

P3a Might Be a Major Factor for Hit Rate. Dev1, Dev2 and Dev6 in the context of city-street noise had lower hit rates than in the context of club noise. In particular, Dev1 and Dev2 showed a noticeable decrease. This may be because the Do chord was harder to discriminate from city-street ambient noise. To better understand the cause, we can analyze Dev1's and Dev2's MMN and P3a. From Figure 4(c)(d), one can clearly see that both the Dev1's and Dev2's P3a amplitudes had a large increase in the city-street context (as compared with the club context). This finding might lead us to presume that P3a was a major factor for hit rate decreasing amid city-street ambient sounds, and imply that an appropriate degree of attentional allocation was important [25], i.e., that *too much* attention shifting may actually cause a lower hit rate. Moreover, both Dev1's and Dev2's MMN amplitudes, as shown in Figure 4(a)(b), were almost the same across the two ambient noises. This further confirms that P3a was a major factor for hit-rate change [25]. A similar finding can also be observed and inferred from Dev3: the P3a amplitudes of Dev3 had no statistically significant difference between the two types of ambient noises, and neither did its hit rates. It should be noted that Dev3 had a larger MMN amplitude amid club ambient noise than amid city-street ambient noise, but this did not affect Dev3's hit rate.

In summary, subjects' behavioral results provide us with a basic understanding of the efficiency of different audio notifications [14, 24], but may not reveal the causes for the differences between them, or their relationship with attention shifting. On the other hand, the MMN and P3a results in the present study may provide information about the brain activity that may help us better analyze the auditory perception of, and attention shifting associated with audio notifications. Therefore, behavioral results should be supplemented with more objective physiological responses, to gain a fuller understanding of the different effects of audio notifications in the context of various ambient noises.

EXPERIMENT 2: AUDIO EVALUATION WITH VISUAL TASK

Experiment 1 demonstrated that MMN and P3a can be used to evaluate audio notifications under more realistic experimental conditions than those have been attempted in the past. Another important factor that interferes with audio perception is the user's cognitive state. In particular, devices generally emit audio notifications to notify a person who is occupied by some tasks. It is therefore important to consider the level of attention shifting and task complexity when evaluating audio notifications. The effectiveness of supplementary notifications will be influenced in part by users' current need to concentrate on something else. Hence, in Experiment 2, we modified the previous auditory oddball paradigm and manipulated subjects' mental workload to explore if and how their auditory perception and involuntary attention shifting would be influenced by their cognitive state. It is hoped that our EEG-based approach can be applied to the usability testing of audio notifications in complex and realistic scenarios, in which subjects' mental workloads include some visual tasks.

Twenty-two subjects (ten females), all of whom had previously participated in Experiment 1, consented to take part in Experiment 2. The settings of Experiment 2 were identical to those of Experiment 1 except that we used the Pairs game to control subjects' workload and attention. To maintain the focus of the study and avoid exhausting the subjects, we used only two deviant stimuli (Dev1 and Dev2) while adopting the same two types of ambient noises as standard stimuli. With this design, we concentrated on investigating how a user's audio perception is affected by their mental workload and the intensity of audio notifications (Figure 1).

Task

We modified the auditory oddball paradigm of Experiment 1 as follows. Instead of watching a video, subjects were required to play a Pairs game. Pairs is a card game in which all the cards are laid face down at the start. The player turns over two cards at a time. If two cards drawn at the same time match, they stay face up; otherwise, they are placed face down. The game can be learned quickly, despite the different levels of mental workload that are required for different levels of difficulty [4]. In this experiment, we varied the difficulty level of the game to control the subjects' mental workload, following Warnock et al. [4] suggestion that playing a Pairs game with 6×4 cards and a 60-second time limit is a useful way to quickly build subjects' mental workload.

Hence, we designed the game with two difficulty levels: one using 6×4 cards and the other using 2×2 cards. All subjects could easily complete the 2×2 version, which was designed to keep subjects engaged in the game but without too much workload. Each subject was trained for 10 minutes before the experiment. EEG recording started 15 sec after a game began. In the task of 2×2 Pairs game task, the subjects had to complete consecutive games continuously. If a subject could not complete a game within 10 sec, a new game would start. This avoided the subjects expending too much mental workload. In the task of 6×4 Pairs game task, subjects had to play at least five games continuously in a testing block and were required to complete each game to the best of their ability. The success rate and completion time were recorded. If the subjects found all the matched pairs in a game or the 60-second time limit expired, the game would immediately restart. We set the length of a testing block as 350 sec, because this experiment only used two deviant stimuli. Subjects took at least five minutes' rest between each testing block. Experiment 2 had four testing blocks for each ambient noise. For each subject, we collected approximately 40 trials for each deviant in each difficulty level.

Results of Experiment 2

Figures 5(a)(b) and 5(c)(d) show the results obtained from a three-way RM-ANOVA analysis (2 ambient noises, 2 deviants, and 2 workloads). Mauchly's test indicated that the sphericity was not violated. With regard to MMN response (Figure 5(a)(b)), whether the subjects' workload was high or low, there were statistically significant differences between Dev1 and Dev2 in both ambient-noise contexts ($p<.05$). With regard to P3a response (Figure 5(c)(d)), there was a significant difference in the workload factor (F=27.03, $p<.001$): larger P3a amplitude was observed in the context of the higher workload level (6×4 Pairs). Except for Dev2 in the context of city-street ambient noise, the P3a amplitudes of Dev1 and Dev2 had statistically significant differences between high workload and low workload. Furthermore, comparing the results of MMN and P3a as shown in Figures 5(a)(b) and 5(c)(d), we also observed that all P3a amplitudes were higher amid the city-street ambient noise, whereas no similar trend was apparent in the MMN-potential results.

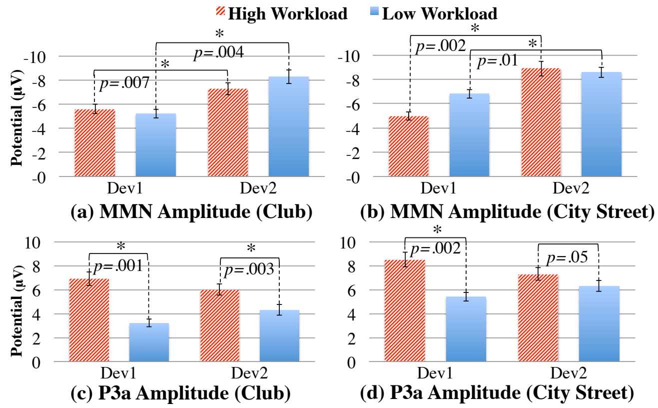


Figure 5. (a) and (b) are MMN amplitudes in the two different ambient noises. (c) and (d) are P3a amplitudes in the two different ambient noises. An asterisk indicates a statistically significant difference ($p<.05$).

Discussion

In this experiment, we modulated subjects' mental workload to investigate its influence on the human perception of, and attention to, audio notifications. Except for Dev2 in the city-street background, we found that the P3a amplitudes of both deviants differed significantly across high and low mental workloads, while MMN amplitudes exhibited no similar phenomenon. This implies that regardless of a user's mental workload, their auditory perception ability is almost the same. This finding is in agreement with those of a prior study [30].

On the other hand, our finding with regard to P3a suggests that subjects allocate more attentional resource to an audio notification when they have higher mental workloads. This finding is inconsistent with earlier research [30]. There are two possible explanations. First, because we used the Pairs game to control subjects' mental workload and focus, audio notifications may have operated more like distractors than notifiers in pervious studies. If a user is concentrating on an assignment, their attention may be involuntary shifted more by obvious audio notifications, and in an extreme case, they might even be frightened by an audio notification if they are too concentrated on a task. Second, the standard stimuli used in this experiment and previous studies [30] are very different; we used ambient noises while previous studies used pure tones. As the context of realistic ambient noise may also influence a subject's cognitive state [19], this should be considered when evaluating the effectiveness of audio notifications.

Taken as a whole, therefore, our findings in experiment 2 suggest us to believe that audio notifications should be designed more carefully, especially when the end-users' mental workload levels are expected to be high. Our results may also suggest that using a louder audio notifications in a higher workload situations might not be useful. From Figure 5(c)(d), we can observe that when the audio intensity is small (Dev1), mental workload has a stronger influence on attention shifting, as its P3a varies more. Conversely, if audio intensity is large (Dev2), workload has a weaker influence upon attention shifting. Although the audio notifications applied in this experiment were irrelevant to the Pairs game, it may be useful in future research to design the task-relevant experiments, e.g., incorporating audio alerts in online gaming scenarios or audio notifications in social media.

EEG-BASED APPROACH TO AUDIO EVALUATION

Comparison to Previous Studies

As the past studies [24, 28] have mentioned, it is difficult to measure and quantify a person's cognitive state during interaction. Although some questionnaires, e.g., NASA-TLX, can be used to investigate a subjects's cognitive state, these surveys are performed off-line after the subject completes a task. As the subject has to recall his/her experience of the experiment, questionnaires may be easily influenced by external factors [24]. Previous studies [30] have shown that users' expectations regarding tones would cause different attention states. Hence, experimental results are probably biased where only reaction time and hit rate have been used to evaluate the effectiveness of notification tones. Our approach allows us to conduct audio evaluation passively, i.e., the subject does not need to pay full attention to the audio notifications to be evaluated. This is consistent with the real-world role of audio notifications, and allows the evaluation procedures to be dovetailed into the real application scenarios.

Our experiments also revealed the impact of sound intensities upon MMN and P3a amplitudes. A prior study [11] pointed out that some notifications alter users to alert quickly, but can also result in higher error rates, especially in environments characterised by high mental demand. This is probably because these audio notifications attract too much attention and lead the users to be distracted from their primary task. We also observed from our experiments that ambient noises can influence MMN and P3a. Therefore, how to keep a balance between auditory perception and attention allocation is an important issue.

Potential Applications of MMN and P3a in HCI

The phenomena of MMN and P3a have been extensively studied within the field of cognitive science. Many studies have applied MMN to clinical practices [16]; however, MMN and P3a have rarely been applied to HCI. We have extended prior work to demonstrate that MMN and P3a can have applications in HCI. Drawing on previous studies on MMN and P3a in cognitive science, we found a number of potential future applications for HCI, and make the following suggestions.

User Experience. Past studies of MMN and P3a have validated many basic physical properties of sound – such as frequency, intensity, and timbre – that are easy to convey and that form memory meanings [22, 23]. Recently, several researchers have begun to test the effects of complex sound stimulation, such as speech sound and composed music (melody, harmony and rhythm). Moreover, subjects' experience of auditory perception has also been investigated, e.g., familiarity with the sounds, learning and/or memory effects [9], and the emotional effect of content [10]. MMN and P3a could help us analyze the impact on audio perception and attention shifting of audio notifications being learned, as well as understand how sound influences users emotionally. MMN and P3a can also measure the changes in users' experience of audio tones.

Adaptive Interface. According to Wetzel et al. [21], who investigated involuntary and voluntary attention shifting on subjects of different ages by P3a. They found that the same auditory stimuli would induce different P3a and cause stronger or weaker distraction in different age groups. Prior work also confirms that trained experts (e.g., musicians) would have the discrepancy with general users on MMN and P3a for the same auditory stimulation. Alcohol and drugs would also influence the users' perception [20]. MMN and P3a have also been studied in single-subject level in recent years. Therefore, MMN and P3a may inspire more diverse ways for the design and evaluation of audio notifications. These findings also suggest that designers should pay more attention to the impact of audio notifications for different user groups.

Context of Use. Depending on the purpose of the notification, designers must also take into account the end-users' likely environments and mental states. Past study [15] on the effect of ambient sounds was mainly conducted such that subjects were focused and prepared to evaluate deviant sounds. However, people process many familiar environmental sounds effortlessly. As such, a change in environments could cause a variety of effects, such as attention being distracted; while ambient sounds that may encourage productivity have also recently been studied [19]. The evaluation of audio notifications when the user is performing a major task is also very important. Whether audio stimuli are relevant or irrelevant to the major task may cause different auditory perception and attention shifting.

The studies [9, 10, 15, 19, 20, 21, 22, 23, 16] mentioned above showed that audio-notification design needs more serious perceptual consideration and EEG-based approaches may be good candidates for its evaluation. Nevertheless, discrepancies inevitably exist between the experimental conditions in cognitive science studies and the real-life situations, so findings may need to be verified or modified if they are to be of use to practical HCI applications.

LIMITATION

There are currently a wide variety of audio notifications, and they are used extensively in technical devices. Despite the diversity of audio notifications, we only chose a small set of them in our experiments, since the goal of this paper is to introduce a new approach for audio evaluation. Recent studies have also extended MMN and P3a to the analysis of emotion, sound learning, memory consolidation, and even the semantic meaning of a sound [20], e.g., whether voices belong to family members or strangers. In our future work, we would like to evaluate more sounds and explore related topics and potential applications.

CONCLUSION

In this study, we have proposed an objective method for evaluating audio notifications. In the first experiment, we verified that MMN and P3a can be used to measure audio notifications in more realistic scenarios. In the second experiment, we demonstrated that this EEG-based approach can be used when the subjects perform major complex tasks, by directly evaluating the changes in users' auditory perception at different levels of workload in the course of an interactive process. Based on the findings of these two experiments, measurement of MMN and P3a can be successfully applied to the usability testing of audio notifications. Moreover, our approach provides a new means by which HCI researchers can investigate the impact of sounds upon individuals' auditory perception and mental states. We believe that this EEG-based approach can open new perspectives for evaluating the human experience of sound.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful comments. This work was supported in part by the National Science Council (101-2628-E-009-021-MY3 and 102-2221-E-009-082-MY3) and the UST-UCSD International Center of Excellence in Advanced Bioengineering sponsored by the Taiwan National Science Council I-RiCE Program under grant number: NSC-102-2911-I-009-101.

REFERENCES

1. Barrass, S., and Frauenberger, C. A communal map of design in auditory display. In *Proc. ICAD* (2009).
2. Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. Earcons and icons: their structure and common design principles. *Hum.-Comput. Interact. 4*, 1 (1989), 11–44.
3. Brewster, S. A. Non-speech auditory output. *Human Computer Interaction Handbook, Lawrence Erlbaum Associates* (2002), 220–239.
4. David, W., Marilyn, M.-L., and Stephen, B. The role of modality in notification performance. *INTERACT* (2011), 572–588.
5. Edworthy, J. Does sound help us to work better with machines? A commentary on rauterberg's paper 'about the importance of auditory alarms during the operation of a plant simulator. *Interacting With Computers 10* (1998), 401–409.
6. Escera, C., Alho, K., Schröger, E., and Winkler, I. Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology 5* (2000), 151–166.

7. Ewing, K., and Fairclough, S. The impact of working memory load on psychophysiological measures of mental effort and motivational disposition. *Human Factors: A system view of human, technology and organisation* (2010), 1–11.

8. Fagerlönn, J., and Liljedahl, M. Awesome sound design tool: a web based utility that invites end users into the audio design process. In *Proc. ICAD* (2009).

9. Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbély, A. A., and Achermann, P. Human central auditory plasticity associated with tone sequence learning. *Learning & Memory 11*, 2 (2004), 162–171.

10. Goydke, K. N., Altenmüller, E., Möller, J., and Münte, T. F. Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research 21*, 3 (2004), 351–359.

11. Graham, R. Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics 42*, 9 (1999), 1233–1248.

12. Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P. N. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *CHI* (2008), 835–844.

13. Helander, M. G., Landauer, T. K., and Prabhu, P. V. Auditory interfaces: Handbook of human computer interface. *Elsevier Science* (1997).

14. Hollands, J. G., and Wickens, C. D. *Engineering psychology and human performance.* Prentice Hall New Jersey, 1999.

15. Kirmse, U., Jacobsen, T., and Schröger, E. Familiarity affects environmental sound processing outside the focus of attention: An event-related potential study. *Clinical neurophysiology 120* (2009), 887–896.

16. Kujala, T., Tervaniemi, M., and Schröger, E. The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology 4* (2007), 1–19.

17. Lemmens, P. M. C., Bussemakers, M. P., and Haan, A. D. Effects of auditory icons and earcons on visual categorization: the bigger picture. *In Proc. ICAD* (2001), 117–125.

18. McGregor, I., Larsson, P., and Turner, P. Evaluating a vehicle auditory display: comparing a designer's expectations with listeners' experiences. In *proc. ECCE* (2011), 89–92.

19. Mehta, R., Zhu, R. J., and Cheema, A. Is noise always bad? Exploring the effects of ambient noise on creative cognition. *JCR 39*, 4 (2012), 784–799.

20. Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology 118*, 12 (2007), 2544 – 2590.

21. Nicole, W., Andreas, W., Stefan, B., and Erich, S. The development of involuntary and voluntary attention from childhood to adulthood: A combined behavioral and event-related potential study. *Clinical Neurophysiology 117* (2006), 2191–2203.

22. Novitski, N., Tervaniemi, M., Huotilainen, M., and Näätänen, R. Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research 20*, 1 (2004), 26–36.

23. Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., and Näätänen, R. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology 118* (2007), 177–185.

24. Peck, E. M., Yuksel, B. F., Ottley, A., Jacob, R. J. K., and Chang, R. Using fNIRS brain sensing to evaluate information visualization interfaces. In *CHI* (2013), 473–482.

25. Polich, J. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology 118*, 10 (2007), 2128–2148.

26. Rafa, A., and Catherine, G. Usability of non-speech sounds in user interfaces. *In Proc. ICAD* (2008).

27. Satu, P., Minna, H., and Risto, N. The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology 121* (2010), 1043–1050.

28. Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., and Jacob, R. Brainput: enhancing interactive systems with streaming fNIRS brain input. In *CHI* (2012), 2193–2202.

29. Stavros, G., Simon, J., Tim, J., and Eamonn, O. Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. *CHI* (2009), 1513–1522.

30. Sussman, E., Winkler, I., and Schröger, E. Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin and Review 10*, 3 (2003), 630–637.

31. Vi, C. T., and Subramanian, S. Detecting error-related negativity for interaction design. *CHI* (2012), 493–502.

32. Winkler, I., Teder, Sälejärvi, W., Horváth, J., Näätänen, R., and Sussman, E. Human auditory cortex tracks task-irrelevant sound sources. *Neuroreport 14*, 16 (2003), 2053–2056.

33. Zander, T. O., and Kothe, C. Towards passive brain–computer interfaces: applying brain–computer interface technology to human-machine systems in general. *J. Neural Eng. 8*, 2 (2011), 025005.