Did You Notice? Neuronal Processing of Multimodal Mobile Phone Feedback

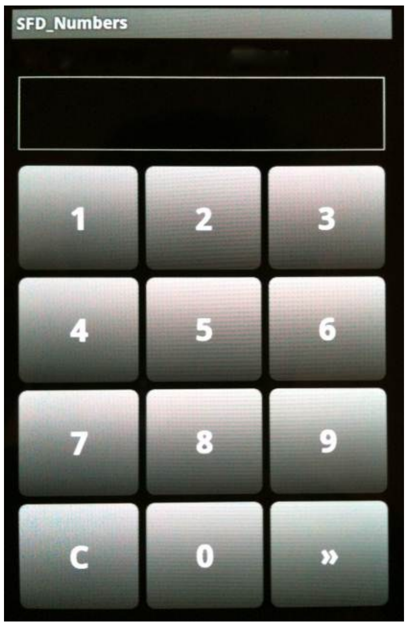

Figure 1 Screenshot of the smartphone display during the test.

Jan-Niklas Antons
Quality and Usability Lab
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
jan-niklas.antons@telekom.de

Sebastian Arndt
Quality and Usability Lab
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
sebastian.arndt@telekom.de

Julia Seebode
Quality and Usability Lab
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
julia.seebode@zmms.tu-berlin.de

Robert Schleicher
Quality and Usability Lab
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
robert.schleicher@telekom.de

Sebastian Möller
Quality and Usability Lab
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
sebastian.moeller@telekom.de

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
To acknowledge information received by a mobile device, a number of feedback modalities are available for which human information processing is still not completely understood. This paper focuses on how different feedback modalities are perceived by users introducing a test method that is new in this field of research. The evaluation is done via standard self-assessment and by analyzing brain activity [electroencephalogram (EEG)]. We conducted an experiment with unimodal and multimodal feedback combinations, and compared behavioral user data to EEG data. We could show that EEG is a feasible method for quantifying conscious processing of feedback in different modalities as it correlates highly with subjective ratings. EEG can thus be considered an additional tool for assessing the effectiveness of feedback, revealing conscious and potential non-conscious information processing.

Author Keywords
Electroencephalography; EEG; multimodal feedback; mobile phones; test method

ACM Classification Keywords
D.2.2 [Software Engineering]: Design Tools and Techniques; H.1.2 [Information Systems]: User/Machine Systems; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems

Introduction

HCI and smartphones

Designing interactions for smartphones is a complex task. Due to the size of the device, standard input devices such as keyboard and mouse have to be replaced by new paradigms, such as gestures on a touchscreen or virtual keyboards. However, in comparison to its physical counterpart, a touchscreen offers no physically-linked feedback to the user. As a consequence, errors or unwanted user inputs can remain unnoticed.

As a replacement for physically-linked feedback, modern mobile phones offer the opportunity of visual, auditory and haptic feedback of any kind, and also any combination of those three. The developer has to decide which option to use in an application. Hence, the question arises how different kinds of feedback are perceived and evaluated by the user. This information could lead to the selection of an optimum feedback strategy. For example, for crucial keystrokes the most informative (noticeable) feedback could be presented to the user, whereas for less crucial keystrokes only minimally-invasive feedback options could be selected.

In an unpublished pre-study, we evaluated subjective user judgments on different feedback modalities when handling a smartphone's touchscreen. We found that different feedback modalities and their combination were not judged alike. Multimodal feedback was subjectively not evaluated better than unimodal. Auditory feedback was judged to be most negative, although this applies only for constant feedback (for each keystroke).

Other studies show that adding a modality to visual feedback is increasing performance for typing [1] [2]. Using solely user performance and /or subjective values it is just passably possible to distinguish between suitable feedbacks types in some scenarios. When it comes to alertness or attention directed towards the last performed action, more detailed results would be desired.

Therefore we deploy physiological data of subjects with the aim to learn more about the subjects' cognitive processing underlying the perception of different feedback types. The underlying rationale is explained in the next section.

Physiological parameters

For measuring cognitive parameters like workload and attention, questionnaires may not cover all relevant facets, e.g. short-term changes, which physiological parameters might help to reveal. One commonly used method to assess neuronal activity is the electroencephalogram (EEG). When using EEG several electrodes are placed on the participant's scalp surface. They measure voltage differences due to neuronal activity in a non-invasive way. Concerning its measurement characteristics, EEG provides a very good temporal resolution and gives the opportunity to conduct tests in a relatively realistic environment compared to methods such as functional magnetic resonance imaging (fMRI), but EEG is limited with regard to spatial resolution.

Voltage differences that are the result of neural activity elicited due to external stimuli are called event-related potentials (ERPs). One of the most prominent features of ERPs is the P300, which is a positive peak occurring usually at around 300 ms after stimulus onset and is related to higher cognitive processes [3] [4]. It varies with stimulus features like intensity or conspicuousness. When a degraded stimulus is presented in a series of undisturbed stimuli and identified by the partici-

	PIN-Digit			
	1.	**2.**	**3.**	**4.**
Block 1	-	-	-	A
Block 2	-	-	-	H
Block 3	-	-	-	V
Block 4	-	-	-	AH
Block 5	-	-	-	AV
Block 6	-	-	-	HV
Block 7	-	-	-	AVH

Table 1 Overview of the test procedure. During each of the seven blocks 80 PINs had to be entered by the participant. 40 of them were without any feedback and the other 40 with the specified feedback on the fourth entered digit. Abbreviations are for the respective presented feedback: A: auditory, H: haptic, V: visual, AH: auditory-haptic, AV: auditory-visual, HV: haptic-visual, AVH: auditory-visual-haptic.

pant, a P300 is elicited. Furthermore, it could be shown that the P300 varies with the level of stimulus degradation for audio [5] and for video [6]. We are interested in ERPs related to mobile phone feedback since they would allow looking at neural processes of phone users that might be partly non-conscious, in contrast to self-assessment such as questionnaires or interviews.

Experimental Preview

In the present study, we investigate which of the available types of feedback attracts most attention of mobile phone users. Therefore, we combine standard questionnaires, measures of reaction time, and physiological data, i.e. EEG. We expect the single visual feedback to attract least attention, and therefore to have the smallest P300 component in the EEG data (Hypothesis 1), since this is one of the standard feedbacks and users are used to it. Furthermore, more attention and therefor a larger P300 component is to be expected whenever haptic feedback is given (Hypothesis 2). Similar results we expect for the questionnaire data. Consequential, we would claim the tri-modal feedback to be most alerting for a participant (Hypothesis 3).

This article is organized as follows: In the next sections a description of our study is given, followed by the results gathered in the subjective test and the EEG measurement. A subsequent discussion of the results paves the way for future work necessary to fully understand the alerting nature of smartphone feedback.

Experiment

Task and stimuli

The experimental setup followed the guidelines of ITU-T Recommendation P.910 for audiovisual experiments in terms of sound isolation and illumination. After giving the participants a description of the experiment and a consent form, they were asked to fill in a demographic questionnaire (items like age, gender). During the experiment the participant’s task was to enter a four-digit personal identification number (PIN) on a standard touchscreen number pad (see **Figure 1**) and to confirm their entry with an additional button press. The PINs were presented on a printed list lying in front of the participants. The order of PINs was randomized for each participant. After entering the last digit of a PIN, they had to press the ’Next’ button at the lower right corner in order to confirm the current PIN. They were asked to input the digits with the index finger of their dominant hand and hold the phone with the other hand. This is important for the measurement of reaction time, since there is a difference in reaction expected when pressing buttons with the dominant hand versus the other.

The smartphone used in this experiment was a Samsung Nexus S with Android 2.3.1 (4.0 inch touchscreen). The experiment was divided into 7 blocks with 80 trials each; 40 trials were without any feedback and the other 40 with feedback. Feedback occurred only after the fourth digit of a PIN to indicate the completion of a PIN. The user was confronted with three single types of feedback, which were visual (V), auditory (A) and haptic (H) plus all possible combinations (AH, AV, HV and AVH) resulting in seven different types of feedback. The feedback type was not changed within a block but rather between them, and the order of blocks was randomized (see **Table 1**). The visual feedback was a switch into the inverted background color of the pressed button for 100 ms. The auditory feedback consisted of a standard Microsoft Windows click sound with 10 ms duration, presented via the integrated loud-

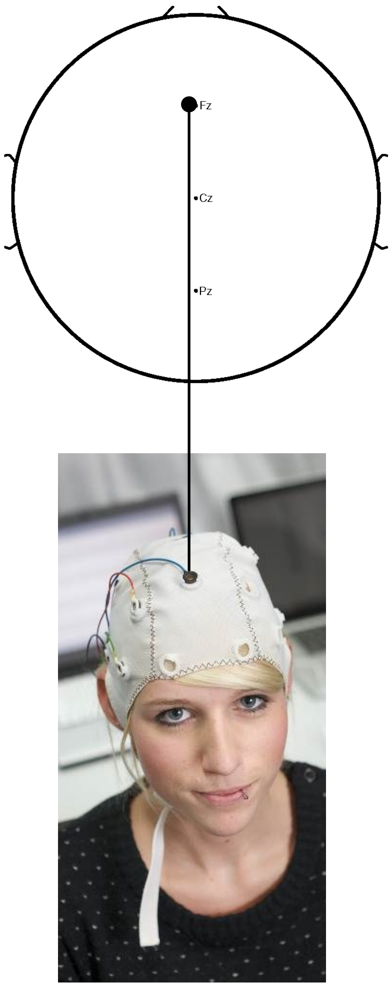

Figure 2 EEG cap with positions measured during the experiment. Nose points upwards. Picture of subject with corresponding electrode marked.

speakers of the smartphone. The tactile feedback was a standard vibration of the phone for about 50 ms. All these types of feedback are commonly used on smartphones. During the test an EEG was recorded. After each block the participants were given a questionnaire, asking for the perception of the feedback just tested during this block. At the end of the test, another questionnaire asking for the likability of all tested feedback types was to be filled in by the participants. The entire test took 2 hours (including breaks).

Methods

Questionnaire

After each block the same questionnaire was provided which served as the subjective test method. It consisted of six items. The items to be tested were: 1) I perceived the feedback consciously. 2) The feedback felt pleasant. 3) The feedback was helpful. 4) The feedback was annoying. 4) The feedback promoted my alertness. 5) The feedback was unpleasant. These had to be rated on the discreet scale: not at all(1) - little(2) - some(3) - quite(4) - entirely(5).

EEG

The EEG (Ag/AgCl electrodes, Brain Products GmbH, Garching, Germany) was mounted with three standard scalp locations according to the extended 10-20 system (Fz, Cz and Pz) (see **Figure 2**). The reference electrode was placed on the tip of the nose; recording was done at a sampling rate of 1000 Hz, and impedances were kept below 10 kOhm. EEG epochs were extracted 100 ms before until 1000 ms after each stimulus (i.e. the fourth entered digit). Peak amplitude and latency were extracted in a fixed time window ranging from 200 ms to 500 ms.

Reaction time

The third parameter measured during the experiment was the reaction time. This value was determined by taking the time difference between the fourth entered digit of each PIN, the potential digit giving feedback, and the hit of the 'Next' button. We extracted whether different feedbacks resulted in different times between PIN completion and confirmation clicking. A Milton-Friedman-Test with post-hoc comparison was calculated to find significant differences. The error rate meaning PIN entered with wrong digits was not analyzed.

Results

Questionnaire

Fifteen healthy German students and university staff (11 male) conducted the experiment. All of them were right-handed with age ranging from 19 - 30 (mean = 24.62 years). None of them reported any hearing or visual impairment or had corrected vision. 80 % of them had experience with touch-controlled mobile devices. For the questionnaire provided after each block, a Milton-Friedman-Test with a post-hoc comparison was calculated. Tri-modal feedback was rated as significantly more annoying. Furthermore, alertness of visual feedback was judged to be significantly lower compared to the rest of the feedback types, except for the auditory feedback. The consciousness rating yielded a significantly higher value for the tri-modal feedback compared to single auditory, visual and the combined auditory-visual feedback which confirms our Hypothesis 3 (all important statistical values, such as the mentioned pairwise comparisons between two modalities (column 2 and 3) and the statistical main effects can be found in **Table 2**).

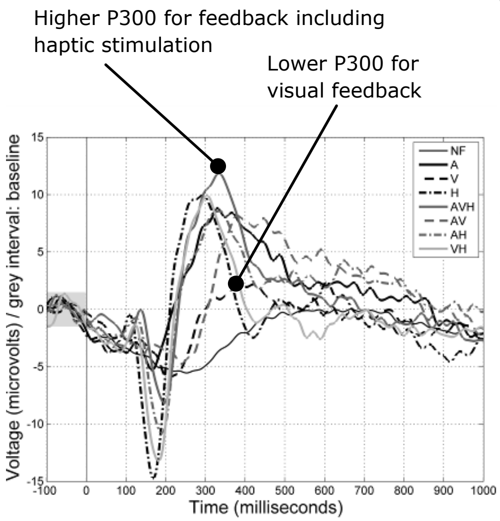

Figure 3 Grand Average ERPs at electrode Cz for all types of feedback, and the no feedback (NF) condition. Feedback occurred after the fourth digit was entered.

EEG

An analysis of variance (ANOVA) with stimulus as a within subject factor revealed a significant main effect for the peak latency ($F(6,48) = 4.10$, $p < .01$, $\eta^2 = .33$) and peak amplitude ($F(6,48) = 2.67$, $p < .05$, $\eta^2 = .25$) of the P300. This partly supports Hypothesis 2. When looking at the Sidak-corrected pair-wise comparisons for peak amplitude, only a significant difference between Auditory-Visual-Haptic and Auditory feedback alone could be found ($p < .05$). For the analysis of peak latency it could be observed that the P300 peak evoked by the haptic feedback is significantly earlier compared with auditory feedback ($p < .05$). The Auditory-Visual-Haptic feedback resulted in the highest P300 peak amplitude. On the other hand, the visual feedback produced the lowest peak of all possible feedbacks, this confirms Hypothesis 1. As can be seen in **Figure 3**, whenever haptics were included in the feedback, a small positive peak from 110 ms to 150 ms followed by a steep trailing edge from 175 ms to 200 ms is arising with a peak around 300 ms. For the peak latency the haptic feedback provided the shortest latency. Additionally, a significantly smaller P300 amplitude for the visual compared to any feedback containing haptic feedback could be observed.

Reaction time

For the measured reaction times a ranking list was compiled and the mean ranks were calculated (main effect for the reaction times ($\chi^2(6) = 42.21$, $p < .01$)). This yielded significantly shorter reaction times for visual (comparison to all other feedbacks $p < .05$) and haptic feedback (comparison to all other feedbacks $p < .05$). The tri-modal feedback resulted in the longest reaction time ($p < .05$), see **Figure 4**.

Discussion and conclusion

The assessment of physiological data during interactions with smartphones is a relatively novel approach in usability testing. It looks promising to be a supplementary method for assessing mental state parameters like alertness. Based on the results of this experiment, we can suggest that EEG is a possible method to access cognitive processing during the interaction with mobile devices. It could thus serve as a tool for interaction designers and usability researchers, revealing information on the early stages of cognitive processing. On the long run we want to use an extended analysis method to be able to reveal potential non-conscious information in more detail as well e.g. the cognitive state of users.

The questionnaire showed that multimodal feedback is drawing more attention to the executed action compared to the haptic and auditory part. With the EEG we could confirm this statement, since we could show that multimodal feedback is resulting in a higher amplitude for the P300 than other types of feedback. Thus, the brain seems to allocate more attention towards multimodal compared to unimodal input. This can also be confirmed with our third type of measurement, the reaction time: again, a significantly higher reaction time for multimodal feedback confirmed the higher attention. This cognitive overload could be used to warn the user before executing a crucial step on their phones. Furthermore, we found that visual-only feedback (which is the most common feedback method used on smartphones) yielded the lowest alertness level across our measurements, and that experienced users considered auditory feedback as the most annoying.

item	feedback	feedback	value
conscious	main		$\chi^2(6) = 20.24, p < .01$
	AVH	A	$p < .05$
		V	$p < .05$
		AV	$p < .05$
	V	H	$p < .05$
		AH	$p < .05$
		VH	$p < .05$
annoying	main		$\chi^2(6) = 15.71, p < .05$
	AVH	A	$p < .05$
		V	$p < .05$
		H	$p < .05$
		AV	$p < .05$
		AH	$p < .05$
		VH	$p < .05$
alertness	main		$\chi^2(6) = 14.86, p < .05$
	V	H	$p < .05$
		AV	$p < .05$
		AH	$p < .05$
		VH	$p < .05$
		AVH	$p < .05$

Table 2 Overview of whether significant values were achieved with rank test for the three most important items and with 'main' describing the main effect for the corresponding item.

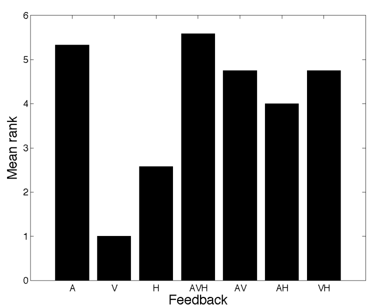

Figure 4 Mean ranks of reaction time for each condition over all participants.

In sum, we could distinguish between several types of feedbacks using performance and subjective indices. ings. These initial results lead us to the conclusion that with an extended analysis facilitating physiological data we will learn more about the underlying processes evoked by the usage of mobile devises.

Future work

The results of our study may lead to the development of an additional tool for designing feedback mechanisms in HCI. An improved test set-up which is more similar to the actual usage situation might be better for interface designers, as it might provide ecologically more valid results. In this study, we always used the no-feedback condition as a baseline / standard stimulus condition to compare with. In common smartphone applications, however, there is usually at least some visual feedback on the phone display, so we might consider using visual-only feedback as a baseline / standard. In our study, we decided to start with a no-feedback baseline to determine some kind of ground truth in ERP differences due to stimulus modalities. In addition to this, standard audio files in the mobile context for clicking could be used in order to get less discrepancy between sound and context. Also real life disturbances like light reflections and ambient noise could be incorporated in the test set-up to examine feedback perception under more realistic conditions. These issues being analyzed in future experiments; we expect that the test methodology presented here might be applicable to the analysis of other interaction issues relevant for mobile phones, such as annoyance and distraction caused by different output modalities, reminders and security alerts.

Acknowledgements

The authors are grateful to Christine Kühnel for providing the test application and Ahmad Abbas for assisting us with the data collection. We would like to thank the Bernstein Focus: Neurotechnology - Berlin (BFNT-B) which supported our work by the Federal Ministry of Education and Research (BMBF) grant FKZ 01GQ0850.

References

[1] Hoggan, E., Crossan, A., Brewster, S., and Kaaresoja, T. Audio or tactile feedback: Which Modality When? In Proc. *CHI 2009*, ACM Press (2009).

[2] Hoggan, E., Brewster, S., and Johnston, J. Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In Proc. *CHI 2008*, ACM Press (2008).

[3] Coles, M.S., and Rugg, M. Event-related brain potentials: an introduction. In *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford University Press, 1995.

[4] Polich, J. Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 10 (2007), 2128-2148.

[5] Antons, J.-N., Schleicher, R., Arndt, S., Möller, S., Porbadnigk, A.K., and Curio, G. Analyzing Speech Quality Perception Using Electroencephalography, *IEEE J. Select. Topics Signal Proc.*, 6, 6 (2012), 721-731.

[6] Arndt, S., Antons, J.-N., Schleicher, R., Möller, S., and Curio, G. Perception of low-quality Videos analyzed by means of Electroencephalography. In Proc. *QoMEX 2012*.