DOES A "NATURAL" SONIC FEEDBACK AFFECT PERCEIVED USABILITY AND EMOTION IN THE CONTEXT OF USE OF AN ATM?

P. Susini IRCAM susini@ircam.fr

N. Misdariis IRCAM misdarii@ircam.fr

O. Houix IRCAM houix@ircam.fr

G. Lemaitre IRCAM lemaitre@ircam.fr

ABSTRACT

The present study[1] examines the question of a "natural" sonic feedback associated with keys of a numerical keyboard - in the context of use of an Automatic Teller Machine (ATM). "Natural" is defined here as an obvious sound feedback with regards to the action made by a user on a device. The aim is then to study how "naturalness" is related to the perceived usability and the perceived emotion of the sonic feedback before and after participants perform several tasks with the keyboard. Three levels of "naturalness" are defined: causal, iconic, and abstract. In addition, two levels of controlled usability of the system are used: a low level and a high one. Results show that pre-experimental ratings of perceived "naturalness" and perceived usability were highly correlated. This relationship held after the participants interacted with the keyboard. "Naturalness" and emotional aspects were less dependant, revealing that "naturalness" and usability represent a special type of relation. However, results are affected by the level of controlled usability of the system. Indeed, the positive change at the high level of controlled usability for the iconic sounds (medium level of naturalness) obtained after the performance task failes at the low level of controlled usability.

1. INTRODUCTION

Designing new artefacts that may be used in everyday life situations reveals several questions. One of those is related to the interaction between the user and the artefact: is a "natural" relation well appropriated to favour an interaction? By natural, it is meant a causal relation instead of an arbitrary one that tends to be perceived as an obvious display of the interaction based on our everyday experiences (referring to the ecological viewpoint by J.J. Gibson [1], a perceived "natural" interaction could be considered as a perceived affordance). In their recent study Rath et al. [2] defined a "natural" interaction as a 'spontaneous understanding of interaction principles on the side of a user'. Within the framework of the present study, we specify the previous question by examining if a "natural" interaction favours the perceived usability of a device? An intuitive answer to this question is provided by D. Norman [3]: "I believe that our reliance on abstract representations and actions is a mistake and that people would be better served if we would return to control through physical objects, to real knobs, sliders, buttons, to simpler, more concrete objects and actions". In the realm of sound perception, recent trends have focused on everyday listening [4, 5] engaging the cause of the sound event rather than selected specific aspects of the sound signal. Briefly, most of the common listeners focus on the cause of the sound, identifying properties of the source (like the size or the material) and the action made by or to the object. Based on the latter assumption, that we are good at identifying sound events, the design of sound for interactive devices has been proposed using a causal display, rather than abstract one. The hypothesis is that "natural" sonic interactions with virtual objects should be perceived as more intuitive. Thus the question in the present study is to test if a "natural" sonic feedback affects the degree of the perceived usability: 1) before interacting with a device, and 2) after users interact with it, in order to examine if the initial impression holds after a period during which the device is used. In addition, relation between "naturalness" and emotion will be examined. Natural sonic feedback was tested by comparison with designed and with arbitrary sonic feedback on a naturalness dimension.

2. EXPERIMENT

The procedure of the present experiment was partly based on the procedure proposed by Tractinsky et al. in [6]. The experiment was a one between-subjects, one within-subjects full factorial design, with the naturalness as the within-subjects factor, and the level of controlled usability as the between-subject factor. Factor 1 was the naturalness of the sonic feedback, with 3 levels (Low / Medium / High). Factor 2 was the controlled usability of the device, with two levels of usability: low and high. For the high level, the sonic feedback operated each time a key was pressed, and for the low level, the sonic feedback did not operate each time a key was pressed

[1] The development of this study was supported by the EC Project CLOSED, FET-NEST no. 29085.

SMC 2009, July 23-25, Porto, Portugal

Participants
Two groups of 45 participants performed the main experiment. Their ages varied from 15 to 58 years old. Each group performed step 1 to 4 in one usability condition. No subject reported having any hearing problems.

Apparatus
The keyboard used was a Mobile Numeric USB Keyboard.

Stimuli
The sound corpus was calibrated along a perceptual scale defining three discrete levels of naturalness: i) *high* – causal – corresponded to various keyboard sound recordings ii) *medium* – iconic or designed - corresponded to synthetic sounds having a causal morphological aspect (impacts) superimposed with non natural timbres iii) *low* – abstract – corresponded to sounds having no relationship with the action made on the keyboard (bicycle ring, piano chord, etc. …)

Procedure
Figure 1 presents the different steps of the present procedure.
Step 0: eighty-one sounds were rated by 20 participants on a scale between the labels "not natural at all" and "very natural" in terms of relation between the sound display and the action of pressing a key on a keyboard. Finally, 9 sounds were selected: three at high level, three at an intermediate one, and three at the lower level of naturalness (see details above)
Step 1: the 9 selected sounds were rated on a nine-points scale by two groups of 45 participants on five scales (Naturalness / N, Usability 1 / U1, Usability 2 / U2, Pleasantness / P and Stimulating / S). U1 and U2 were described to the participants as scales related to the a priori perceived usability of the sound (before using the keyboard in step 2). For example, for scale Usability 1 / U1, participants were asked to rate the assertion: "I find that the sound is well associated with the keys". P and S were associated to the two usual emotional dimensions used in the realm of research in emotions [7]. For example, for scale Pleasantness / P, participants were asked to rate the assertion: "I find that the sound is pleasant". For each scale, the mark 1 was associated with "I don't agree at all" and the mark 9 with "I completely agree".
Step 2: based on their evaluations in step 1, each participant was assigned at one of the three level of "naturalness" (H/M/L), and then performed step 2 using one sound that she/he evaluated at the corresponding level of "naturalness". In step 2, using the numerical keyboard, different tasks were performed several times like withdrawing cash and transferring an amount of money between two bank accounts.
Step 3: the sound used during the performance task was rated on the same five scales as in step 1.
Step 4: participants were asked to rate directly if the sound was finally worse or better compared to their initial impression on each scale, which corresponds to the direct difference between post and pre-experimental perception. For example, for the pleasantness, the assertion proposed was: "I find that the sound is *more* pleasant". Evaluations were made on the scale [-4, +4]; a negative mark meant that the participant did not agree, and a positive one that she/he agreed.

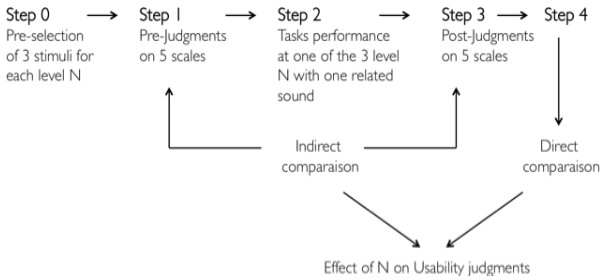

Figure 1. Step 0 to 4 of the experimental procedure

3. RESULTS

3.1 Manipulation check

Figure 2.a and 2.b displays the pre-experimental mean ratings for the naturalness scale in order to check that the 9 sounds were correctly judged with the expected level of naturalness. The mean value for each sound corresponds to the average of the 45 evaluations on this scale respectively for group 1 and group 2. As it can be observed on figures 2.a and 2.b, pre-selected sounds from step 0 were judged in step 1 with the same level of naturalness as it was expected. On the figures, the first three sounds are labelled respectively H1, H2 and H3 as they were expected to be perceived with the highest level of naturalness, the three next sounds are labelled respectively M1, M2 and M3 for the medium level of naturalness, and the last three ones are labelled L1, L2 and L3 respectively for the lowest level of naturalness. A repeated-measures analysis of variance (ANOVA) was performed on pre-experimental ratings obtained on the naturalness scale with one within-subject factor, sound (9 levels), and with one between-subject factor, group of participants (2 levels). The analysis reveals a strong effect of the sound factor (F(8,

704)=167.8, p < 0.001) and no effect of the group factor, as well as no interaction between the two main factors which means that ratings did not depend on the group factor. As it was expected that the first three sounds [H1, H2, H3] will be judged with a higher level of naturalness compared to [M1, M2, M3] and [L1, L2, L3], and [M1, M2, M3] to be judged with a higher compared to [L1, L2, L3], contrast analyses were performed in order to test if perceived naturalness ratings were different between the three groups of sounds. Results showed that perceived naturalness was significantly different between [H1, H2, H3] and the two other groups of sounds (p<0.001), and between [M1, M2, M3] and [L1, L2, L3] (p<0.001)

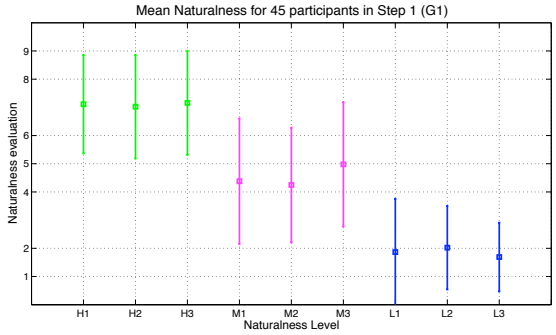

Figure 2.a Naturalness pre-experimental mean ratings on 45 participants from Group 1

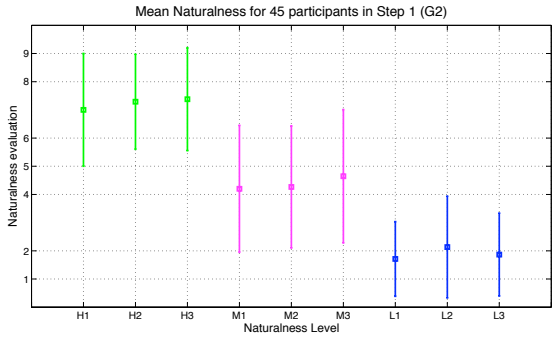

Figure 2.b Naturalness pre-experimental mean ratings on 45 participants from Group 2

3.2 Analysis of the pre-experimental ratings

3.2.1 Reliability

For the two groups of subjects, the repetition (test and retest) factor is examined considering individual ratings on each of the five scales. Test-retest reliability obtained is r=0.83, r=0.78, r=0.71, r=0.69 and r=0.60 (p < 0.001) respectively for the five scales (Naturalness / N, Usability 1 / U1, Usability 2 / U2, Pleasantness / P and Stimulating / S). Since the correlation values for several scales were not very high, even if they were statistically significant, datasets were not aggregated. In addition, participants reported to have been more confident in their second ratings. Based on participants' comments, only the ratings from the second set were kept for further analysis.

3.2.2 Results presentation

Figure 3.a and 3.b displays the pre-experimental mean ratings of the five scales (Naturalness, Usability 1, Usability 2, Pleasantness, Stimulating) for respectively group 1 and 2 (high and low usability level). The mean values for each level of naturalness for the five scales are also presented in Table 1.

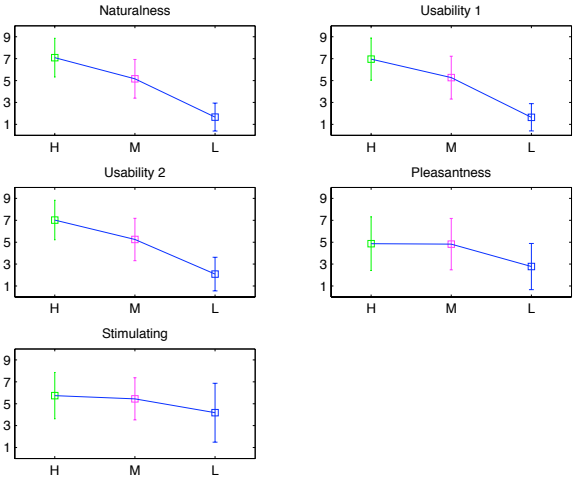

Figure 3.a Ratings on the five scales for the sounds assigned to one of the three level of naturalness (High, Medium, Low) and for the **high** level of the controlled usability.

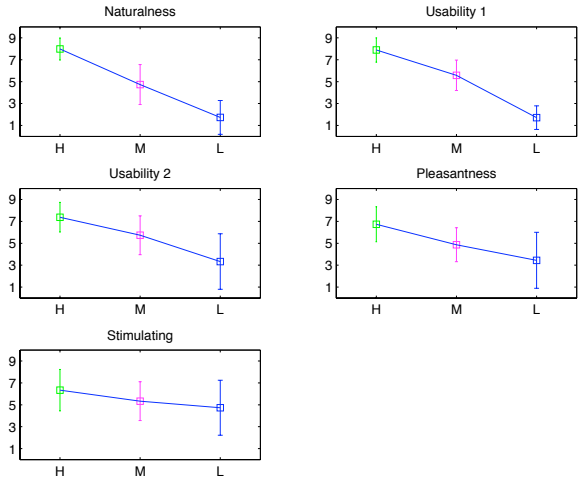

Figure 3.b Ratings on the five scales for the sounds assigned to one of the three level of naturalness (High, Medium, Low) and for the **low** level of the controlled usability.

Naturalness level	Pre-experimental scale	Controlled Usability	
		High	Low
High	Naturalness	7.08 (1.76)	7.97 (1.01)
	Usability 1	6.95 (1.93)	7.88 (1.11)
	Usability 2	7.02 (1.80)	7.37 (1.35)
	Pleasantness	4.86 (2.45)	6.73 (1.60)
	Stimulation	5.73 (2.11)	6.33 (1.89)
	N	*15*	*15*
Medium	Naturalness	5.15 (1.77)	4.73 (1.82)
	Usability 1	5.26 (1.95)	5.57 (1.38)
	Usability 2	5.24 (1.93)	5.73 (1.77)
	Pleasantness	4.82 (2.34)	4.86 (1.54)
	Stimulation	5.44 (1.92)	5.33 (1.77)
	N	*15*	*15*
Low	Naturalness	1.66 (1.27)	1.73 (1.54)
	Usability 1	1.64 (1.24)	1.71 (1.07)
	Usability 2	2.08 (1.53)	3.33 (2.53)
	Pleasantness	2.77 (2.10)	3.44 (2.55)
	Stimulation	4.17 (2.69)	4.73 (2.50)
	N	*15*	*15*

Table 1. Pre-experimental mean ratings (standard deviations in brackets) on the five scales and for both level of usability

Results show clearly that the degree of perceived usability U1 and U2 decreased with the same amount as the perceived naturalness N, whereas the amount of change of the perceived pleasantness P is slightly less important. Finally, the level of stimulation (S) seems independent of the three level of naturalness. This result indicates that participants perceived a stronger relation between naturalness and usability rather than between naturalness and emotional aspects.

3.2.3 Analysis of variance

A multivariate analysis of variance (MANOVA) with repeated measures was conducted on the five dependent variables (Naturalness, Usability 1, Usability 2, Pleasantness, Stimulating), using a full-factorial design, with the following two between-subject factors: level of naturalness (3 levels, H, M and L) and group of participants (2 levels, G1 and G2). A MANOVA was performed instead of an ANOVA to take into account correlations between ratings on similar scales such as U1 (Usability 1) and U2 (Usability 2), for example.
The main interest here is to determine whether the naturalness factor has had a global effect on ratings for the five scales. Results show, as it was expected, that the factor group neither had an effect on ratings nor interacted with the factor naturalness. On the other hand, the multivariate analysis of variance (MANOVA) reveals an overall significant effect of the naturalness factor (Wilks' lambda value, $F=32.0$, $p<0.001$). One-way ANOVAs show that the effect is significant for each scale. The percentage of total variance accounted for by each effect is indicated by the R^2 coefficient. The main effect of naturalness accounts for about 81, 58, 32 and only 15% of the total variance respectively for scales U1, U2, P, and S. Thus the strongest effect of the naturalness factor is obtained for ratings on U1 and U2. This result corroborates descriptions provided in the previous section.

3.3 Comparison between Pre and Post-experimental ratings

3.3.1 Correlation analysis

Inter-correlations among the perceived measures are presented in table 2.a and 2.b. Pre experimental ratings of perceived Naturalness and perceived Usability 1&2 were highly correlated (respectively $r=0.9$ and $r=0.8$). The same results were obtained for Post experimental ratings (respectively $r=0.85$ and $r=0.71$) meaning that the correlations between perceived Naturalness and Usability 1&2 remained high even after the performance task. On the other hand, Pre-experimental ratings of perceived Naturalness were less correlated with the scale Pleasantness ($r=0.62$) and weakly with the scale Stimulating ($r=0.44$). The weakest correlations were obtained for ratings on the Stimulating scale and the other scales for both Pre and Post experimental ratings (respectively $0.44 \leq r \leq 0.56$ and $0.36 \leq r \leq 0.42$). This indicates that the level of correlation with the Naturalness scale depends on the type of scale. Pre and Post-experimental correlations of Naturalness were relatively high ($r=0.78$), and Pre and Post-experimental correlation of perceived Usability 1&2 were lower (respectively $r=0.65$ and $r=0.55$), as well as Pre and Post-experimental correlation for scales Pleasantness and Stimulating (respectively $r=0.34$ and $r=0.41$). There was a slightly difference between the two usability groups related to the perceived Pleasantness; the Pre and Post-experimental correlation were 0.39 and 0.29 respectively for the high and low-usability groups.

	Pre-U1	Pre-U2	Post-N	Post-U1	Post-U2
Pre-N	0.90***	0.80***	0.78***	0.69***	0.54***
Pre-U1		0.79***	0.72***	0.65***	0.52***
Pre-U2			0.65***	0.59***	0.55***
Post-N				0.85***	0.71***
Post-U1					0.71***

Table 2.a Correlation matrix of pre and post-experimental measures (N=90) for the Naturalness (N), the Usability 1 (U1) and the Usability 2 (U2) scales (***p<0.0001, **p<0.005, *p<0.05)

	Pre-P	Pre-S	Post-N	Post-P	Post-S
Pre-N	0.62***	0.44***	0.78***	0.54***	0.32**
Pre-P		0.56***	0.47***	0.34**	0.28**
Pre-S			0.28*	0.23*	0.41**
Pots-N				0.72***	0.36**
Post-P					0.42***

Table 2.b Correlation matrix of pre and post-experimental measures (N=90) for the Naturalness (N), Pleasantness (P) and Stimulating (S) scales (***p<0.0001, **p<0.005, *p<0.05)

3.3.2 Results presentation

Figure 4.a and 4.b display the average ratings for the two groups of participants (respectively the two levels of usability) obtained in step 4 showing how each level of naturalness was perceived on the five scales before and after interacting with the keyboard in step 2. A positive value indicates that positive change in ratings between before and after the performance task.

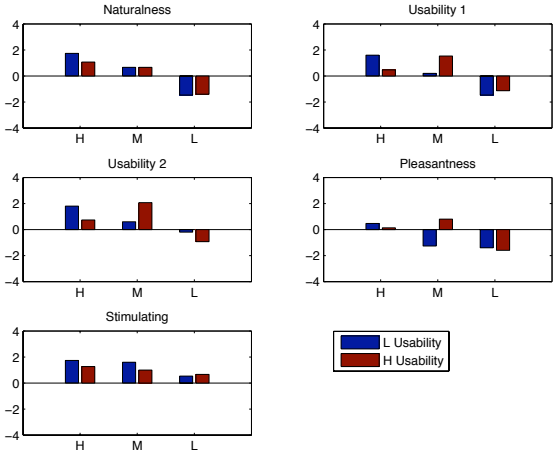

Figure 4. Direct estimation of the difference between post and pre-experimental perception on the five scales and for both level of usability

Analysis of variance

A multivariate analysis of variance (MANOVA) with repeated measures was conducted on the direct estimation of the difference between post and pre-experimental perception on the five scales. The main interest here is to compare ratings obtained for the two usability levels in order to examine effect of the controlled usability on perceptive ratings. Thus, the main null hypothesis tested is: "the experimental condition, controlled usability, does not have any effect on perceptive ratings". The MANOVA reveals an overall significant effect of the naturalness factor (Wilks' lambda value, $F=4.42$, $p<0.001$) but no effect of the controlled usability factor. On the other hand, the analysis reveals a significant interaction between these two factors. It thus appears that the effect of the controlled usability (between-subjects factor) is present but it varies as a function of naturalness (within-subjects factor). One-way ANOVAs reveal that the naturalness factor is significant for all the scales except for the scale Stimulating. In addition, the analyses reveal that the significant interaction is obtained only for perceived Usability 1&2 ($F(2, 84)=15.7$, $p<0.001$ and $F(2, 84)=7.8$, $p<0.001$, respectively) and for Pleasantness ($F(2, 84)=6.64$, $p<0.01$). Contrast analyses show that the interactions obtained for these three scales are based on significant difference only for the medium level (M) of Naturalness, while there is no significant difference between the two other levels of Naturalness. These analyses reveal that the controlled usability affect only the medium level of naturalness that corresponds to the iconic (designed) sounds. As it can be seen in figure 4, for the low level of controlled usability (L Usability), iconic sounds were judged after the performance task to be less usable and pleasant.

4. CONCLUSION

This study examined whether a “natural” sonic feedback affects the degree of the perceived usability and emotion in a simple interaction. Results show that:

- the naturalness scale related to the three groups of sounds was perceived as expected in the context of use of the keyboard.
- initial impression of naturalness affects the initial impression of usability and holds after participants interact with the keyboard. On the other hand, ratings on the emotional scales are less affected.
- the level of controlled usability interacts mainly with the medium level of naturalness. Sounds at this level were designed sounds (synthetic impact sounds), well perceived prior to the performance task (pre-judgment) and finally perceived to be more useful in a high level of usability, but this impression failed in the low usability condition.
- whatever the situation, an abstract sound, corresponding to a low level of naturalness, is not perceived to be useful and pleasant, and even worse after performing with the keyboard.

To summarize the major findings of this study, the results suggest that it looks like the acceptance of an iconic (designed) sound is weaker when the sound is not efficient for the expected feedback

(dysfunctioning system). However, a causal sound is still accepted when the system does not work correctly.

5. REFERENCES

[1] Gibson, J. J. (**1966**). *The Senses Considered as Perceptual Systems* (Houghtin Mifflin, Boston).

[2] Rath, M., and Schleicher, R. (**2008**). "On the Relevance of Auditory Feedback for Quality of Control in a Balancing Task," Acta Acustica united with Acustia 94, 12-20.

[3] Norman, D. A. (**1999**). " Affordance, conventions, and design," Interactions 6 6, 38-43.

[4] Gaver, W. W. (**1993**). "How do we hear in the world? Explorations in ecological acoustics," Ecological Psychology 5, 285-313.

[5] Gaver, W. W. (**1993**). "What do we hear in the world? An ecological approach to auditory event perception," Ecological Psychology 5, 1-29.

[6] Tractinsky, N., Adi, S.-K., and Ikar, D. (**2000**). "What is Beautiful is Usable," Interacting with Computers 13, 127-145.

[7] Russell J. A., “A circumplex model of affect”, Journal of personality and social psychology, vol. 39(6), pp. 1161-1178, 1980.