J Multimodal User Interfaces
DOI 10.1007/s12193-011-0086-0

ORIGINAL PAPER

Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback

Patrick Susini · Nicolas Misdariis · Guillaume Lemaitre · Olivier Houix

Received: 17 January 2011 / Accepted: 23 December 2011

Abstract This study examined the influence of the naturalness of a sonic feedback on the perceived usability and pleasantness of the sounds used in a human-computer interface. The interface was the keyboard of an Automatic Teller Machine. The naturalness of the feedback was manipulated by using different kinds of relationship between a keystroke and its sonic feedback: causal, iconic, and arbitrary. Users were required to rate the naturalness, usability, and pleasantness of the sounds before and after manipulating the interface. Two kinds of interfaces were used: a normally functioning and a defective interface. The results indicated that the different relationships resulted in different levels of naturalness: causal mappings resulted in sounds perceived as natural, and arbitrary mappings in sounds perceived as non-natural, regardless of whether the sounds were recorded or synthesized. Before the subjects manipulated the interface, they rated the natural sounds as more pleasant and useful than the non-natural sounds. Manipulating the interface exaggerated these judgments for the causal and arbitrary mappings. The feedback sounds ruled by an iconic relationship between the user's gesture and the resulting sounds were overall positively rated, but were sensitive to a potential contamination by the negative feelings created by a defective interface.

Keywords Sonic interactions · Interfaces · Naturalness · Usability · Pleasantness

G. Lemaitre is now at the University IUAV of Venice, Italy, research unit interactions.

P. Susini (✉) · N. Misdariis · G. Lemaitre · O. Houix
IRCAM, Paris, France
e-mail: susini@ircam.fr

1 Introduction

Sound is widely used in variety of products, ranging from desktop computers to mobile phones, and from coffee makers to electric cars. One of the most important issue for the sound designers is that of the relationships between the sounds and the information to be conveyed to the users.

Two kinds of information are usually considered in the design process. The first one is related to the function of the sound, as regards the use of an object or an interface. The aim in this case is to design sound characteristics that will establish an efficient interaction between the user and an object or an interface: for example, reaction to an alarm sound is faster when the alarm sound has a rapidly repeating pulse. The second kind of information is related to the global coherence between a sound and an object or the identity of a brand. In this case, the aim can be for example to design a pleasant sound.

1.1 Auditory displays: mapping sounds and information

A variety of interfaces commonly use sounds to display information to a user. Some applications have used sounds to let users explore large amounts of complex data (e.g. seismic data [6]). More commonly, sounds in interfaces provide a feedback to a user's action (e.g. the deletion of a computer file), or warn the person that something is happening (e.g. an alarm clock) [13].

In fact, the design of warning signals has became increasingly sophisticated, in particular for applications that need to display different warnings with different meanings and different urgency levels to the users. Such interfaces are now relatively common in applications like hospital or car equipment, or high performance aircraft [2, 21, 28]. The design of warning signals provides the designers with an interesting

Published online: 27 January 2012

framework to analyze how sounds can display information. Three types of relationships are usually considered: symbolic (arbitrary relationship), iconic (representational relationship), or causal [16]. Symbolic relationships (e.g. different tone frequencies coding different levels of urgencies, or the Morse code) allow the designer to map any kind of information to any kind of sound parameter, but requires the users to learn the mapping. Iconic (e.g. a downward pitch to signal the loss of altitude to the pilot of an aircraft) and causal relationships (e.g. the squeaking sound of a car braking to signal a potential accident to car driver) limit the design possibilities to existing referential sounds, but relies on relationships already learned by the users.

Sounds in human-computer interfaces can be analyzed in the same framework. Historically, designers of human-computer interfaces have focused on sounds in the form of short abstract static signals, typically warning or feedback sounds. The iconic and arbitrary relationships have been used to respectively design two types of sounds for human-computer interactions: auditory icons and earcons. Auditory icons are iconic caricatures of sound occurring in everyday life situations that are transposed in a virtual environment [7]. Earcons are arbitrary simple sounds that are related to elementary actions, in a virtual environment, using hierarchical rules for creating more complex messages [1].

The most preeminent example of an auditory icon is probably the sound of a sheet of paper being crumpled and thrown down to a garbage can, used as a feedback to file deletion. This sound was originally designed as a part for *Sonic Finder*, a suite of sounds developed for Apple Computers, Inc. [8, 9]. Another example of a feedback created in Sonic Finder was a scraping sound when the user dragged a file on the desktop. In this latter example, the causal relationship between the sound and its meaning was created by using the sound *naturally* resulting from a user's gesture (at least metaphorically): dragging an object on a surface.

In fact, sonic feedback, both natural and created, are quite common in interactions with everyday objects. They are important because they tell the user that his or her action has resulted in the expected consequence. For instance, door latches make a characteristic impact sound when they correctly catch in the socket. This sound indicates that the door is securely locked, and a silent door lock would be utterly inconvenient. The design of auditory icons, or more generally sounds based on a causal relationship between a sound and its meaning, therefore relies on the fact that users have already integrated a number of relationships between a mechanical event and its sonic consequences: listeners spontaneously identify the cause of a sound [17]. The meaning of interfaces based on causal relationships are expected to be intuitively understood by users.[1]

1.2 Causality and naturalness

The perception of causality is closely tied to the synchrony between a user's gesture and the resulting sound [11]. For example, whether two moving discs with crossing trajectories are perceived as bouncing or overlapping is heavily affected by the presence, timing and nature of a sound occurring at the contact instant [10].

But even more important are the expectations formed by an individual interacting with his or her environment. The sounds that a user "commonly" expects as a result of his or her gesture (i.e. the sound of an impact resulting from striking an object) are here referred as "natural". The natural relationships between a sound and a gesture are those driven by the laws of physics [20].

1.3 Sonic interactions: mapping sounds and actions

Causal relationships have been mainly studied for continuous *sonic interactions* [25]. In fact, the tight coupling between action and sounds has lead designers to create sonic interfaces that involve the active and continuous manipulation by the users [15]. Such interfaces are, for instance, used to let users explore complex sonified data by actively manipulating sound sources and processings [14]. Other examples are the sonification of an athlete's movements to help him or her achieve better performance [3, 26], or the design of pseudo-haptic interfaces in wearable computers [5]. Even though our research was concerned with discrete interactions,[2] it is first important to examine the knowledge that can be gathered from such studies.

Two recent examples are worth detailing. In the first example, Lemaitre and colleagues [19] designed a tangible interface (the Spinotron) based on the metaphor of a child's spinning top: when the users pumped the Spinotron, they drove a physical model of a ratcheted wheel that produced a characteristic clickety-clack sound. The participants were required to pump the interface and to reach and maintain a precise and constant pace (indicated by a visual target). Half of the participants were provided with the continuous auditory feedback, half with only the visual target. Only the participants who were provided with the auditory feedback were able to improve their performance across trials. However, when asked to describe their appraisal of the sonic feedback, the subjects reported two interesting comments: first, they were not aware that the sound actually helped

[1]Note that using a natural or causal relationship may have its own drawbacks—e.g. users having an overly deterministic vision of the feedback model based on prior expectations from the "natural" situation at play.

[2]A *discrete* feedback occurs for a finite and short amount of time as the result of a user's action (think of the beeps of a microwave oven); on the contrary, a *continuos* feedback follows the dynamics of a gesture sustained by the user. The sound produced by a musician bowing a string is a typical example of a continuous interaction.

them improve their performance. Second, they found the sound irritating.

The methodology developed by Rath and colleagues [22, 23] provides another interesting example of a sonic interaction. The Ballancer is a tangible interface consisting of a wooden plank that can be tilted by its user, in order to roll a virtual ball along a plank. A user tilting the latter heard the rolling sound produced by the virtual ball. The authors used this interface to study subjects' abilities to use the auditory feedback in a task involving guiding the ball to a target region along the length of the plank. In a comparison using the same task with the rolling ball sound and with a "synthetic" sound (i.e. one that did not mimic any physical system) that preserved the same information, subjects were found to perform better, early in training, with the causal sound than with the "synthetic" one, whereas the latter provided better performance after training. Participants also reported having preferred the causal sounds.

These studies therefore suggest that using a causal mapping between the user's action and the resulting sonic feedback has two advantages. First, as in the case of passive auditory displays, causal relationships are already integrated by the subjects, and the interfaces are easier to learn. Second, users seem to generally prefer natural over synthetic sounds. It seems likely that sounds causally mapped to the user's actions would be perceived more natural than arbitrarily associated sounds.

However, these studies all focused on *continuous* interactions. The current research studied a type of interaction that is more commonly found in interfaces: a *discrete* feedback.

1.4 How to evaluate an interface?

These studies also point us towards two aspects that are important for the evaluation of interfaces: the usability and the aesthetics of the design.

The usability of an interface is generally indexed by measuring the performances of users required to fulfill a task with the interface. Aesthetics is most of the times considered along two aspects: sensory pleasantness and annoyance. Sensory pleasantness, a notion that directly comes from the psychoacoustical tradition, is considered as an auditory attribute dependent on other auditory attributes such as roughness, sharpness, tonality and loudness [4, 30]. Annoyance is a concept used to describe the nuisance caused by noises, particularly in the case of urban noises. Noise annoyance can be related to acoustic variables, but acoustic characteristics do not play the most important role. Psycho-sociological variables are important determinants [12].

In fact, both the objective usability of an interface and the pleasantness of the sound feedback can potentially influence the user's emotional reactions. According to Scherer's appraisal theory of emotions, both the appraisal of how well the interface allows the user reaching a goal and the pleasantness of the sounds may influence the valence of the user's feelings [27]. Lemaitre and colleagues [18] showed for instance that in the case of an interface that they designed (the Flops), the valence of the user's feelings were independently influenced by the difficulty of the manipulation of the interface and the pleasantness of the sounds (in this case, natural sounds were found more pleasant than artificially irritating ones).

But the usability of an interface is not only a matter of objective performance measurement: the usability of an interface is also *perceived* by the user. The perceived usability of an interface depends on the context, the goal and the skills of the user. It is also influenced by the aesthetics of the interface. Tractinsky et al. [29] showed for instance that the user's initial aesthetic perception of an interface (an Automatic Teller Machine) influenced the perception of the usability, after its manipulation. The objective usability did not have such an effect. To put it in a few words, attractive products are perceived easier to use, or to borrow Tractinsky's words, "What is beautiful is usable".

1.5 Goals of the study

Our goal was to evaluate the influence of sonic feedback on the appraisals of the interface. Specifically, we have investigated how naturalness interplays with the user ratings on pleasantness and perceived usability of the interface.

We used an interface that required one of the simplest gestural interactions possible: pressing a key. Natural sounds could therefore be easily defined as those naturally caused by a keystroke.

Similarly to Tractinsky et al. [29], the keyboard interface was used to interact with an Automatic Teller Machine Interface (ATM). This is a commonly used interface. Moreover, sounds are in this case not a simple embellishment: they help the users correctly enter a Personal Identification Number (PIN). In fact, it is rather common that the sun's reflections on the visual displays makes the visual feedback (a star displayed after each key press) illegible, and the interaction particularly difficult and frustrating. A sonic feedback solves this issue.

More precisely, we were interested in three questions:

- Do the different mappings between a keystroke and the resulting sound change the perceived naturalness of the sound?
- Does the naturalness of the sonic feedback influence the perceived usability and pleasantness of the interface?
- Is the influence of the naturalness affected by the active manipulation of the interface?

This latter question is particularly relevant from a design point of view. Usually, auditory displays are evaluated by having listeners passively listening and evaluating the sounds. However, this underestimates the influence of the manipulation. Lemaitre and colleagues [18] have shown for instance that evaluating sounds alone gives an overemphasized view of the actual influence of sounds on users' feelings when the sounds are embedded in an interactive interface.

Therefore, the study consisted of several experiments in which the subjects interacted with the interface, and the naturalness of the sound feedback was manipulated. The naturalness was manipulated by using three different mappings between the action of pressing a key and the resulting sounds. The highest degree of naturalness was achieved by using a *causal* mapping: the actual sounds of a keystroke. The lowest degree of naturalness was achieved by using sounds that did not bear any semantic relationships with a keystroke (e.g. a bicycle bell): an *arbitrary* mapping. In between, the medium level of naturalness was created by using sounds that had the temporal envelope of keystroke sounds, but a different timbre: an *iconic* mapping.

1.6 Outline

The experimental design was inspired by Tractinsky et al. [29]. It consisted of five steps. In the first step, the participants were presented with the interface and required to rate the naturalness, usability and pleasantness of the sounds. In a second step, the participants were required to perform a number of bank operations by manipulating the interface. In the third and fourth step, they had to indicate how the manipulation of the interface had changed their appraisal of the sounds. The fifth step was used to validate some results observed in the previous steps.

Before reporting on this experiment, the next section will first detail the design of the interface and the sounds.

2 Design of the sonified ATM interface

2.1 The ATM interface

We designed a simplified interface for a bank Automatic Teller Machine (ATM). The interface consisted of a numerical keypad (Mobile Numeric USB Keyboard) connected to a laptop computer. The interface was sonified: when a user hit a key, a sonic feedback was played. The interface was controlled by Cycling 74's Max/MSP v.4.6. The users could use the ATM interface to execute different bank operations. There were two scenarios:

– Withdrawing cash from an bank account. This scenario required the user to enter a bank account number and the amount of money to be withdrawn. The account number consisted of the same four-digit number for all participants.

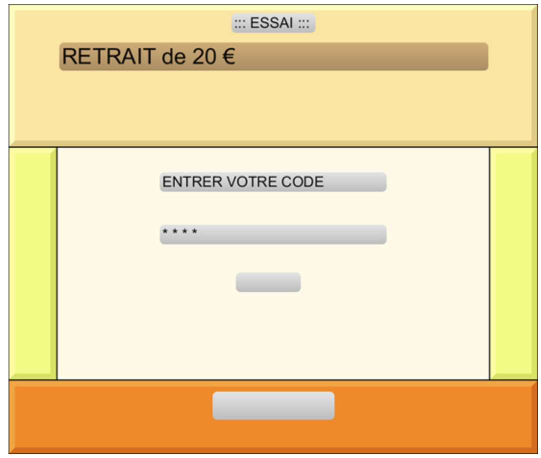

Fig. 1 The graphical interface of the ATM used in the study. The interface was programmed in Max/MSP

– Transferring cash from an account to another. This scenario required the user to enter two bank account numbers and the amount of money to be transfered.

Importantly, the users had to hit the "I" key several times to indicate an amount of money. For instance, they had to press six times the key to indicate an amount of 60 €. A star displayed after each keystroke was used as a visual feedback. This rather inconvenient procedure made the sonic feedback particularly useful for the participants.[3]

The interface had two modes of operation. In the *normal* mode the sonic feedback was systematically played immediately after a key was pressed (normal). In the *defective* mode of operation keystrokes would sometimes fail to trigger the sonic feedback. The probability that the keystroke triggered a feedback was 70%. The graphical of the ATM is represented in Fig. 1.

2.2 Creation of the sonic feedback types

The first experimental step (this step will be labeled "step 0" in the following of the article) was the selection of different sonic feedback types to be associated to a keystroke. Eighty-one feedback sounds were first chosen or created according to three different mappings between the action of striking a key and the resulting sounds (respectively: arbitrary, iconic, and causal). These mappings were intended to create three levels of naturalness (respectively: low, medium, and high).

– **Causal mapping:** Thirty-nine keyboard sounds were selected to create a high level of naturalness. These sounds were sampled from The SoundIdeas' General 6000 library.[4]

[3]Note that this interface was only developed for the sake of the experiment. We do not advise to use such an interface in a real ATM.

[4]http://www.sound-ideas.com/6000.html.

Table 1 The different sounds used to sonify the ATM keyboard

Label	Origin	Description	Duration (ms)
H1	Recording	Computer keystroke (LF)	203
H2	Recording	Computer keystroke (MF)	241
H3	Recording	Computer keystroke (HF)	315
M1	Synthesis	Similar to a metallic ball	176
M2	Synthesis	Similar to a wooden clave	237
M3	Synthesis	Synthetic HF impulse	499
L1	Recording	Bicycle bell	527
L2	Recording	Spring-like sound (Foley)	737
L3	Recording	Fast ascending movement on a piano keyboard	998

HF: high-frequency; MF: medium-frequency; LF: low-frequency

- **Iconic mapping:** To create a medium level of naturalness, 21 sounds were created by cross-synthesis[5] [24]: an impulse envelope was applied to non-natural timbres. This procedure allowed us to create non-natural sounds with the same morphology as the sound of an impact.
- **Arbitrary mapping:** To create a low level of naturalness, 21 sounds were selected, which bore no morphological similarity nor semantic relationship with an impact or keystroke. These sounds did not have an impulsive time envelope, and could not be associated with a keystroke (e.g. bicycle ring, piano chord, etc.).

2.3 Selection of the feedback sounds

To select the sounds that best conveyed the three levels of naturalness, a set of participants was required to rate the naturalness of the sounds.

Before rating the sounds, they saw a video illustrating the three different levels of naturalness. The video displayed a hand holding a hammer and striking a floor made of ceramic tiles, once. The video was presented three times, each time with a different soundtrack. The first sound was the real sound of the impact on the floor (causal mapping); the second was an artificial impulse sound (iconic mapping); the third one was a synthetic chord (arbitrary mapping). The participants were instructed that these three videos illustrated examples of a high, medium and low level of naturalness.

Twenty participants were required to rate the naturalness of the 81 sounds. They used a three-point scale with the following labels: "Not natural at all", "In between", and "Very natural".

[5] We used the cross-synthesis module for Max/MSP developed at Ircam http://imtr.ircam.fr/imtr/Max/MSP_externals.

From these ratings, we selected three sounds for each of the three levels of naturalness (making a total of nine sounds). For each level, we selected three sounds that more than 90% of the participants had rated at this level. In following of the article, the three levels of naturalness will be labelled L (low), M (medium) and H (high), and the sounds will be labeled L1, L2, L3 (lowest level of naturalness), M1, M2, M3 (medium level of naturalness), and H1, H2, and H3 (highest level of naturalness). The nine sounds are listed in Table 1 and are available at http://pds.ircam.fr/atm_sounds.html.

3 Rating the sonic feedback

To assess how the different levels of naturalness influenced the user's evaluation of the pleasantness and usability of the feedback sounds, we conducted an experiment. The procedure of this experiment had several steps that are detailed in the next paragraphs. At each of these steps, the participants were required to rate different feedback sounds on different scales: naturalness, usability, and pleasantness. We manipulated two experimental factors expected to influence the ratings: the *naturalness* of the sonic feedback (factor 1, with three levels: low, medium and high), and the *mode of operation* of the interface (factor 2, with two levels: normal and defective).

3.1 Three scales of evaluation

The participants were required to rate the different feedback sounds associated to the keys of the ATM, before (step 1: pre-experimental ratings) and after using the ATM (step 2: manipulation; step 3: post experimental ratings). For each of these scales, they were presented with an assertion (see below), and had to qualify the assertion by using a nine-point scale ranging from "I do not agree at all" (1) to "I completely agree" (9). These scales were:

Naturalness (N): The scale of naturalness was defined similarly as in step 0: the participants had to assess wether they found that the sound was a natural consequence of the action of striking a key. This scale was described by the assertion: "I find that the sound is well associated with the keys".

Usability (U) For the scale of usability, the participants had to rate how useful they found the sounds for typing in the different operations of the ATM. This scale was defined by the assertion: "I find that the sound helps me type in information". Note that in step 1 (see below), the users did not manipulate the interface. Their ratings of usability were therefore only based on their listening to the sounds.

Pleasantness (P) For this scale, the participants had to rate how pleasant they found the sounds. They had to qualify the assertion: "I find this sound pleasant".

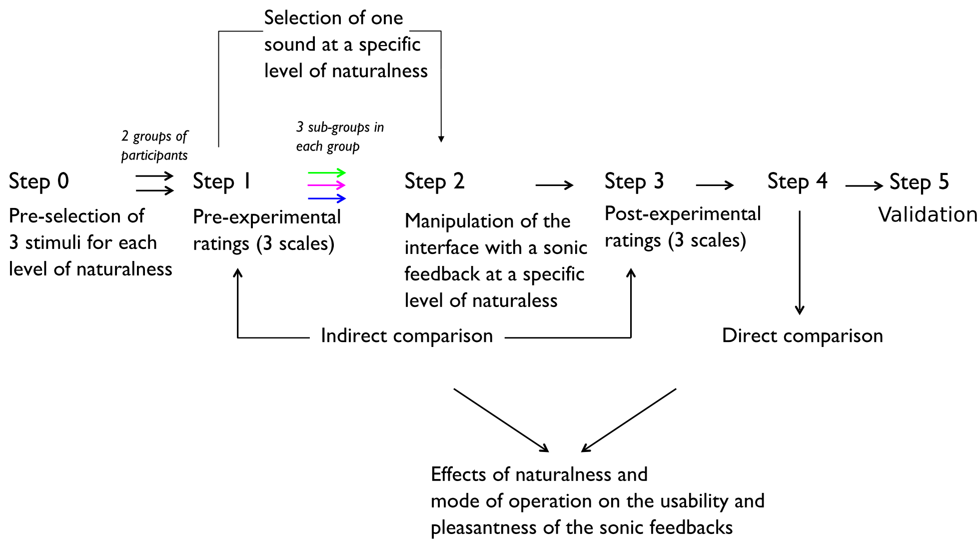

Fig. 2 Different steps of the experimental procedure

3.2 Method

Participants There were 90 participants (60 females and 30 males) between the ages of 15 and 58. All participants reported normal hearing and were native French speakers. None of them had participated in step 0.

Apparatus The experiment was run on an Apple MacBook Pro laptop computer (Mac OSX 10.5), using Cycling 74's Max/MSP 5. The sounds were played through the computer built-in loudspeakers.

Stimuli The nine stimuli described above were used in the experiment.

Procedure The participants were split in two groups. Half of them used the normal mode of operation of the ATM (Group 1), and the other half used the defective mode of operation (Group 2). The procedure had five steps that are represented in Fig. 2. These steps were:

Step 1: Pre-experimental ratings. The participants were first presented with the interface and the nine sounds, but could not use the interface. They had to rate the sounds on the three nine-points scale previously described (N, U, and P). Each sound was rated twice on each scale (test-retest).

Step 2: Manipulating the interface Each participant was randomly assigned to one of three sub-groups (L, M, and H), each corresponding to one of the three levels of naturalness (L, M, and H). In each sub-group, the subjects manipulated the ATM sonified with one sound of the corresponding naturalness level. The sounds were selected on the basis of each participant's judgement in step 1: for example, if a participant was assigned to the group L, he or she used a sound that he or she had rated with a low level of naturalness in step 1. This procedure made certain that the participant interacted with the system using a sound that he or she had evaluated at the corresponding level of naturalness. There were 15 participants in each sub-group. The participants performed the previously described tasks (withdrawing cash and transferring money between accounts) several times

Step 3: Post-experimental ratings Each subject rated the sound used in step 2 on the same three scales as in step 1.

Step 4: Direct comparisons The participants were finally asked to directly rate if the sound was worse or better than what they initially thought. This is, therefore, a direct comparison of their pre- and post-experimental assessments. Step 4 was used to prevent a potential ceiling effect in step 3 (for instance, a sound already maximally rated in step 1 could not received a higher rating in step 3). The participants used a numerical scale ranging from -4 to 4 and were presented with an assertion of the type: "I find that the sound was more natural/helpful/pleasant after experimenting with the interface". A negative rating meant that the participant did not agree, and a positive one that he or she agreed. In the following analyses, only the ratings measured in step 4 will be analyzed in detail.

Step 5: Validation The last step was used to validate some of the results observed in step 3 and 4 for the high and medium levels of naturalness by having participants rate the usability and pleasantness of the interface at both levels of naturalness in the same session.

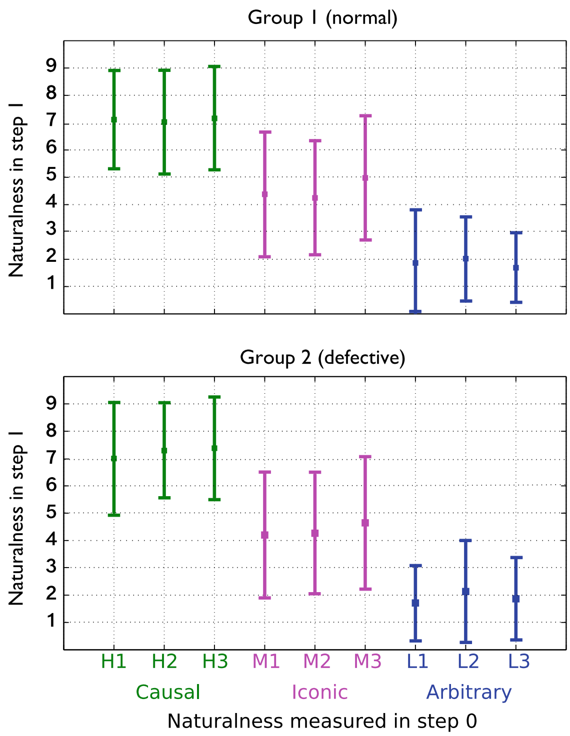

Fig. 3 Pre-experimental ratings of naturalness (N, step 1), averaged across the 45 participants in Group 1 (*upper panel*, normal mode of operation) and Group 2 (*lower panel*, defective mode of operation). The *vertical bars* represent the standard deviation

3.3 Results

3.3.1 Ratings of naturalness in steps 0 and 1

The pre-experimental ratings of naturalness (N) allowed us to verify that the sounds selected in step 0 created the appropriate levels of naturalness for the participants in this experiment (steps 1 to 4). Figure 3 represents these ratings averaged across the 45 subjects in Groups 1 (upper panel) and 2 (lower panel). In both these panels, the horizontal axis represents the naturalness measured in step 0 for the nine sounds, and the vertical axis represents the naturalness measured in step 1. This figure shows that the ratings of naturalness in step 1 preserved the rank ordering of naturalness measured in step 0: the sounds H1 to H3 were rated as more natural than M1 to M3, which were rated in turn as more natural than L1 to L3.

These observations were confirmed by submitting the data to a repeated-measure analysis of variance (ANOVA) with one within-subject factor (the nine sounds), and one between-subject factor (the two groups). The analysis revealed a strong effect of the sounds ($F(8, 704) = 167.8$, $p < 0.001$), no significant effect of the groups, as well as no interaction between the groups and the sounds: the ratings were similar in the two groups. An analysis of contrasts showed that the ratings of naturalness for the sounds H1, H2, and H3 were significantly larger than the ratings of the six other sounds ($p < 0.001$), and that the ratings of M1, M2, and M3 were significantly larger than the sounds L1, L2, and L3 ($p < 0.001$).

Table 2 Pre-experimental ratings for the subjects in the two groups (G1: normal mode of operation; G2: defective mode of operation) on the three scales: average and standard deviation (in brackets)

Naturalness level	Pre-exp. scale	Mode of operation	
		Normal (G1)	Defective (G2)
High	Naturalness	7.08 (1.76)	7.97 (1.01)
	Usability	7.02 (1.80)	7.37 (1.35)
	Pleasantness	4.86 (2.45)	6.73 (1.60)
	N	15	15
Medium	Naturalness	5.15 (1.77)	4.73 (1.82)
	Usability	5.24 (1.93)	5.73 (1.77)
	Pleasantness	4.82 (2.34)	4.86 (1.54)
	N	15	15
Low	Naturalness	1.66 (1.27)	1.73 (1.54)
	Usability	2.08 (1.53)	3.33 (2.53)
	Pleasantness	2.77 (2.10)	3.44 (2.55)
	N	15	15

3.3.2 Analysis of the pre-experimental ratings (step 1)

Test-retest reliability For the two groups of participants, we examined the test-retest reliability by computing the coefficient of correlation between the two measures for each scale. The coefficient of correlations were respectively $r = 0.83$, $r = 0.71$, and $r = 0.69$ ($N = 180$, $p < 0.001$ in the three cases) for the three scales (N, U, and P).

The coefficients of correlation were statistically different from 0, but not very large. Furthermore, the participants reported that they were more confident in their second ratings. Therefore, we only used the second ratings (retest) in the subsequent analyses.

Raw results Figure 4 displays the pre-experimental ratings on the three scales, averaged across the participants in Groups 1 and 2. The average values for the three scales are also presented in Table 2 for each level of naturalness.

The figure and the table first show that the naturalness of the sounds had an influence on the three rating scales. It clearly shows that the ratings of naturalness (N) decreased (between the high and low levels of naturalness) with approximately the same amount (5.42 in Group 1, 6.24 in Group 2) as the ratings of usability (U; 4.94 in Group 1, 4.04 in Group 2), whereas the decrease of the ratings of pleasantness (P) was slightly less important, especially in Group 1 (2.09 in Group 1, 3.29 in Group 2).

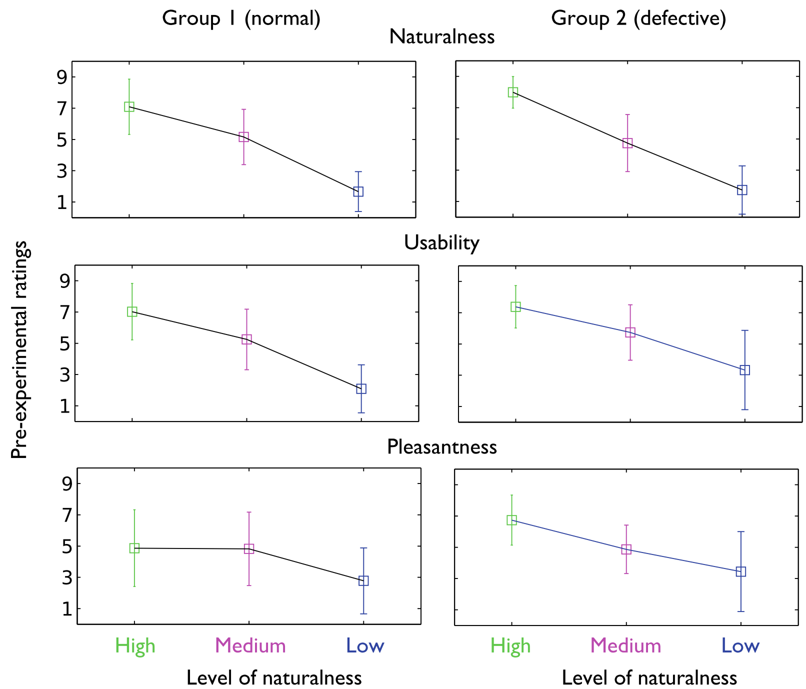

Fig. 4 Pre-experimental ratings on the three scales for the sounds at the three level of naturalness (High, Medium, Low), for the normal mode of operation (Group 1, *left panel*), and for the defective mode of operation (Group 2, *right panel*)

Analysis of variance A multivariate analysis of variance (MANOVA) with repeated measures was conducted on the three dependent variables (N, U and P), using a one between-subject (the two groups), and one within-subject (the naturalness of the sonic feedback with 3 levels) full factorial design. A MANOVA was performed to take into account the potential correlation between the rating scales. The main interest here was to determine whether the naturalness factor globally affected on the ratings on the three scales.

The results of the MANOVA showed that the naturalness of the sounds was the only significant effect (Wilks' lambda value, $F = 32.0$, $p < 0.001$). Three two-way ANOVAs showed that the effect of the level of naturalness was significant for each scale. The percentage of total variance accounted for by each effect was indicated by the R^2 coefficients. The main effect of naturalness accounted for 58 and 32% of the total variance respectively for the scales U and P. Thus the strongest effect of the naturalness factor was obtained for the ratings on U.

3.3.3 Comparison between the pre- and post-experimental ratings (step 3)

The correlation between the pre- and post-experimental (Table 3) ratings of naturalness was relatively high ($r = 0.78$). This indicates that the ratings of naturalness were rather similar before and after the manipulation of the interface. The correlation between the pre- and post-experimental ratings of usability was lower ($r = 0.55$), as well as the correlation

Table 3 Correlation matrix of pre- and post-experimental ratings ($N = 90$) for the naturalness (N), the usability (U), and pleasantness (P) scales. ***: $p < 0.0001$, **: $p < 0.01$

	Post-exp. *N*	Post-exp. *U*	Post-exp. *P*
Pre-exp. *N*	.78***	.54***	.54***
Pre-exp. *U*	.65***	.55***	.44***
Pre-exp. *P*	.47***	.37***	.34**

between the pre- and post-experimental ratings of pleasantness ($r = 0.34$). There was a slight difference between the two groups of participants for the ratings of pleasantness: the correlations between the pre- and post-experimental ratings were respectively 0.39 and 0.29 for Groups 1 (normal mode of operation) and 2 (defective mode of operation). Overall, these correlations suggest that the manipulation of the interface modified the user's appraisal of the sounds' usability and pleasantness. More details are provided by the analysis of the direct comparisons in step 4.

3.3.4 Direct comparisons (step 4)

Results presentation Figure 5 represents the average ratings in step 4 (direct comparisons) for the two groups of participants (the two modes of operation). This figure shows how the participants assessed that the manipulation of the interface had changed their assessments of the sounds on the three scales. A positive value indicates an increase on a particular scale after the manipulation.

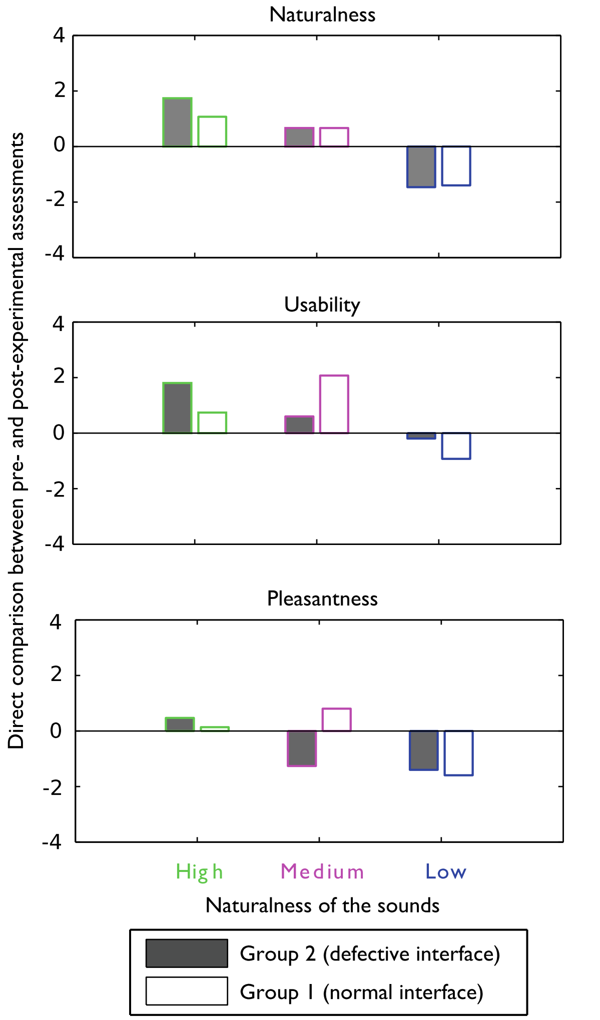

Fig. 5 Direct comparisons of the pre- and post-experimental assessments on the five scales for the two groups

Analysis of variance A multivariate analysis of variance (MANOVA) was conducted on the direct estimation of the difference between pre- and post-experimental assessment of the scales. The main interest was to compare the ratings obtained for the two groups, to examine the effect of the mode of operation (Factor 2) on the ratings of usability. The MANOVA revealed an overall significant effect of the naturalness (Wilks' lambda value, $F = 4.42$, $p < 0.001$) but no effect of the mode of operation. However, the interaction between the two factors was significant: the influence of the naturalness of the sounds on the ratings was modulated by the mode of operation of the interface.

Three ANOVAs revealed that the naturalness of the sonic feedback had a significant effect on the three scales. First, the manipulation of the interface exaggerated the assessment of the naturalness of the sounds (upper panel of Fig. 5): the sounds at the highest level of naturalness were judged more natural after the experiment than before, and the sounds at the lowest level were judged less natural after the experiment. Second, the sounds at the highest and medium level of naturalness were overall judged more usable after the experiment than before the experiment (middle panel of Fig. 5). The sounds at the lowest level of naturalness were judged equivalently usable before and after the experiment. Third, the sounds at the lowest level of naturalness were judged less pleasant after the experiment (lower panel of Fig. 5).

In addition, the analyses revealed a significant interaction between the naturalness of the sonic feedback and the mode of operation of the interface for the scales U ($F(2,84) = 7.8$, $p < 0.001$) and P ($F(2,84) = 6.64$, $p < 0.01$). A set of contrast analyses showed that the interactions were only statistically significant for the medium level (M) of naturalness (there was no significant difference between the two other levels of naturalness): the mode of operation of the interface affected the ratings of usability and pleasantness only for the sounds at the medium level of naturalness (iconic mapping between the gesture and the sonic feedback). When the interface operated normally, the sounds with a medium level of naturalness were considered much more usable after the manipulation than before. This was not the case when the system was defective (see Fig. 5). At the medium level of naturalness, the sounds were considered more pleasant after the experiment than before for the participants using the normally functioning interface. However, the participants using the defective interface judged the sounds much less pleasant after the manipulation.

3.3.5 Validation (step 5)

A final step was done in order to confirm, within the same session, previous results concerning the difference obtained for the high and the medium levels of naturalness. In fact, during the previous steps, results for the high and the medium levels of naturalness were obtained in different sessions in which participants manipulated the ATM with only one sound; each participant was assigned to a group corresponding to one of the three levels of naturalness, and differences were compared among the group of participants.

Results clearly showed that the sounds at the lowest level of naturalness (arbitrary mapping) were judged less pleasant and usable after the manipulation of the ATM. More surprisingly, results showed that ratings of usability and pleasantness were affected for the medium and high levels of naturalness by the mode of operation of the ATM; for the normal mode of the ATM, both levels were judged with more or less the same level of pleasantness, but for the defective mode of operation the medium level was judged significantly less pleasant. For the normal mode of the ATM, ratings of usability were higher for the medium level of naturalness than for the high one, and the opposite result was obtained for the defective mode. Thus, in order to confirm the effect of the mode of operation on the medium and high levels of

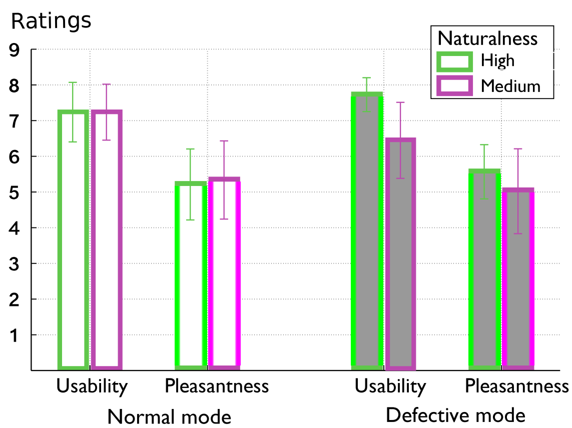

Fig. 6 Step 5. Ratings on the two scales (Usability, Pleasantness) for the sounds at the two levels of naturalness (High, Medium), and for the normal and the defective mode of operations. The *vertical bars* represent the 95% confidence interval

naturalness, one sound at the highest level of naturalness (respectively causal mapping) and a second sound at the medium level of naturalness (respectively iconic mapping) were compared during the same session by each participant. This is the major difference with previous steps.

Thus, in this final step, twenty participants performed the same task as in step 2. Ten started with one sound corresponding to the high level of naturalness and were asked to evaluate the sound on two scales (respectively U and P), then they repeated the same procedure with the sound corresponding to the medium level of naturalness. Among the ten participants, five did the experiment with the normal mode of operation first, and then with the defective one; five did the opposite. The ten other participants started with the medium level of naturalness first.

The results are represented in Fig. 6. For the normal mode, results did not show any difference between the two levels of naturalness for both scales (resp. U and P), whereas, for the defective mode, results reveal a difference for both scales, but the effect was not significant. Only the interaction between the mode of operation and the level of naturalness for the scale of usability (U) was significant ($F(1,19) = 8.19$, $p < .05$). This highlights that both levels of naturalness were rated with a similar level of usability for the normal mode, but not for the defective mode; the sound with the higher level of naturalness was judged more usable and the sound with medium level of naturalness is judged less usable. Even if the effect was not strong, this validation experiment confirmed the effect of the mode of operation on the medium and high levels of naturalness for the scale of usability.

3.4 Discussion

The first conclusion of the experiment is that level of naturalness of the sounds had an important effect on the appraisal of the sounds. In step 1, before the users had a chance to manipulate the ATM interface (pre-experimental ratings), the sounds at the highest level of naturalness were rated as being more usable and more pleasant than the sounds at the lowest level of naturalness. This indicates that the participants a priori better appraised the causal than the arbitrary sounds.

The ratings measured after the participants had manipulated the interface in step 2 (post-experimental ratings in step 3 and direct comparisons in step 4) further qualified this interpretation. An interesting remark is that there was overall no significant difference between the two groups of participants. The mode of operation of the interface had no principal effect on the ratings of the sounds. This suggests that the participants were actually rating the sounds, and not the interface.

However, the manipulation of interface changed the user's appraisal of the sounds. The participants consistently indicated in step 4 that their appraisal had changed, but the changes depended on the levels of naturalness of the sounds. Overall, the sounds at the highest level of naturalness were rated as more natural and more usable after the manipulation than before the manipulation. The sounds at the lowest level of naturalness were rated as less usable and less pleasant after the manipulation than before the manipulation. Therefore, it can be concluded that the ratings in step 4 tended to exaggerate the pre-experimental ratings.

Several explanations can be proposed. First, the participants' longer exposition to the sounds during the manipulation of the interface might have emphasized their initial judgements. However, a closer look at the interaction between the mode of operation of the interface and the level of naturalness of the sounds shed light on an interesting phenomenon, and allows us to propose a second explanation. For the sounds with a medium level of naturalness (the sounds that bore an iconic relationship the action of striking a key), the effect of the manipulation was different for the two groups of participants. The participants in group 1, who used a normally operating interface rated these sounds as more usable and more pleasant after the experiment (similarly to the sounds with high level of naturalness). But the participants in group 2, who used a defective interface rated these sounds almost equivalently usable after and before the experiment, and much less pleasant than what they initially thought. Therefore, this suggests that the exaggeration of the appraisal of the sounds was caused by the experience of the sounds in situation. For the sounds with the highest and lowest level of naturalness (i.e. the causal and arbitrary sounds), the manipulation of the interface confirmed the participants' initial appraisal. But the sounds with a medium level of naturalness (the iconic sounds) became less pleasant when the interface was defective. In other words, the appraisal of the sounds was contaminated by the negative appraisal of the interface: the iconic sounds were only tolerated when the interface worked properly. When the frustrating manipulation of the defective interface emotionally upset the users, they found the iconic sounds irritating and less pleasant. This effect did not occur for the causal (which were consistently positively rated) and the abstract sounds (consistently negatively rated).

4 Conclusion

This study examined three questions. The first question focused on the influence of different kinds of mappings between a keystroke and the resulting sound on the perceived naturalness of the sounds. The results of the experiments clearly indicates that a causal relationship resulted in sounds perceived as natural, whereas an arbitrary relationship resulted in sounds perceived as not natural. By manipulating the relationship between the user gesture required during the interaction and the sounds played as feedback, one can control the perceived naturalness of the feedback sounds.

The second question examined whether the naturalness of the sonic feedback had an effect on the pleasantness and perceived usability of the sounds. The results of the experiments also clearly show that this was the case: natural sounds were perceived as being more pleasant and more helpful than non-natural sounds.

Overall, the durations of the sounds (see Table 1) with a high and medium level of naturalness were similar (because the sounds at the medium level were created with the same temporal morphology as the sounds at the high level). In contrast, the sounds at the low level of naturalness were selected so as to be semantically and morphologically different from a keystroke. Accordingly, the durations for the low level of naturalness were longer. Therefore, the semantic and temporal mismatches are confounded in this case, and the negative effect of the low level of naturalness may be explained by either one or both of these aspects.

The third question examined the effect of manipulating the interface on the ratings of naturalness, pleasantness, and usability. Overall, manipulating the interface exaggerated the user's appraisal of the interface. After the manipulation, the users found the natural sounds more natural and more usable than before the manipulation. They perceived the natural sounds as more natural and usable than they had before the manipulation.

Two modes of operation were used during the experiment: a normal mode, and a defective mode. Comparing the results for these two modes of operation offers an interesting conclusion for the sounds with a medium level of naturalness sounds. These sounds were iconic. They were synthetic sounds that had the temporal envelope of an actual keystroke sound, but a different timbre. Overall, these sounds were rated halfway between the natural and the non-natural sounds in terms of usability and pleasantness. However, the manipulation of the interface had a different effect on the ratings of these sounds, depending on whether the interface was properly working or not. In particular, the iconic sounds were rated as being slightly more useful after the manipulation of the normal interface, as was the case with the natural sounds. But after the manipulation of the defective interface, they were rated as less usable, an effect opposite to the effect of the natural sounds.

These results provide sound designers with some interesting remarks. First, manipulating the mapping between the user's gesture and the sounds changes the perception of the naturalness of the sounds. It is in particular important to note that naturalness was not influenced by whether the sounds were synthetic or recorded sounds: in fact, the less natural sounds were recordings. Sounds that are perceived as the natural consequence of the user's gesture are especially desirable in user interfaces; they are judged as pleasant and useful. The synthetic iconic sounds used in the experiment were rated as more pleasant and useful than the non-natural recordings (even though less pleasant and useful that the non-natural sounds). As such, they offer an interesting alternative when technical reasons prevents the use of recordings. However, our results show they are more susceptible to be affected by the difficulty of the manipulation. The sounds were contaminated by the negative feelings resulting from the manipulation of a defective interface. In other words, these sounds received the blame for the non-working interface, which was not the case when using natural sounds as feedback.

Finally, these results emphasize that designers must test the quality of their sound design by having users actively manipulate the interface. Even with our simplified interface, our results make it clear that the manipulation of the interface greatly influences the appraisal of the sounds.

References

1. Blattner MM, Sumikawa DA, Greenberg RM (1989) Earcons and icons: their structure and common design principles. Hum-Comput Interact 4:11–44
2. Edworthy J, Loxley S, Dennis I (1991) Improving auditory warning design: relationship between warning sound parameters and perceived urgency. Hum Factors 33(2):205–231
3. Eriksson M, Bresin R (2010) Improving running mechanisms by use of interactive sonification. In: Proceedings of the 3rd interactive sonification workshop, Stockholm, Sweden
4. Fastl H (1997) The psychoacoustics of sound-quality evaluation. Acust United Acta Acust 83:754–764
5. Fernström M, Brazil E, Bannon L (2005) HCI design and interactive sonification for fingers and ears. IEEE Multimed 12(2):36–44
6. Fröhlich B, Barrass S, Zehner B, Plate J, Göbel M (1999) Exploring geo-scientific data in virtual environments. In: Proceedings of the IEEE conference on visualization (VIS99), San Fransisco, CA

7. Gaver WW (1986) Auditory icons: using sound in computer interfaces. Hum-Comput Interact 2(2):167–177
8. Gaver WW (1989) The sonic finder: an interface that use auditory icons. Hum-Comput Interact 4:67–94
9. Gaver WW (1994) Using and creating auditory icons. In: Kramer G (ed) Auditory display: sonification, audification and auditory interfaces. Westview Press, New York
10. Grassi M, Casco C (2010) Audiovisual bounce-inducing effect: when sound congruence affects grouping in vision. Atten Percept Psychophys 72(2):378
11. Guski R, Troje N (2003) Audiovisual phenomenal causality. Percept Psychophys 65(5):789–800
12. Guski R, Felscher-Suhr I, Schuemer R (1999) The concept of noise annoyance: how international experts see it. J Sound Vib 223(4):513–527
13. Hermann T (2008) Taxonomy and definitions for sonification and auditory display. In: Susini P, Warusfel O (eds) Proceedings 14th international conference on auditory display (ICAD 2008). Institut de Recherche et de Coordination Acoustique Musique, Paris
14. Hermann T, Bovermann T, Riedenklau E, Ritter H (2007) Tangible computing for interactive sonification of multivariate data. In: Proceedings of the 2nd international workshop on interactive sonification, York, UK
15. Hermann T, Hunt A (2005) An introduction to interactive sonification. IEEE Multimed 12(2):20–24
16. Jekosch U (1999) Meaning in the context of sound quality assessment. Acust United Acta Acust 85:681–684
17. Lemaitre G, Houix O, Misdariis N, Susini P (2010) Listener expertise and sound identification influence the categorization of environmental sounds. J Exp Psychol, Appl 16(1):16–32
18. Lemaitre G, Houix O, Susini P, Visell Y, Franinović K (2012) Emotions are influenced by auditory feedback from a computationally augmented artifact. Trans Affect Comput (in press)
19. Lemaitre G, Houix O, Visell Y, Franinović K, Misdariis N, Susini P (2009) Toward the design and evaluation of continuous sound in tangible interfaces: the Spinotron. Int J Hum-Comput Stud 67:976–993 special issue on Sonic Interaction Design
20. McIntyre J, Zago M, Berthoz A, Lacquaniti F (2002) Does the brain model Newton's laws? Nat Neurosci 4:693–694
21. Patterson RD, Edworthy J, Shailer MJ, Lower M, Wheeler P (1986) Alarm sounds for medical equipment in intensive care areas and operating theatres. Institute of Sound and Vibration Research Report No. AC598, Southampton, UK
22. Rath M, Rocchesso D (2005) Continuous sonic feedback from a rolling ball. IEEE Multimed 12(2):60–69
23. Rath M, Schleicher DR (2008) On the relevance of auditory feedback for quality of control in a balancing task. Acta Acust United Acust 94:12–20
24. Roads C, Strawn J (eds) (1987) Foundations of computer music. The MIT Press, Cambridge. ISBN 0-262-68051-3
25. Rocchesso D, Polotti P (2008) Designing continuous multisensory interaction. In: Contribution to the SID workshop, CHI, Florence, Italy
26. Schaffert N, Mattes K, Effenberg AO (2010) Listening to the boat motion: acoustic information for elite rowers. In: Proceedings of the 3rd Interactive sonification workshop, Stockholm, Sweden
27. Scherer KR, Dan ES, Flykt A (2006) What determines a feeling's position in affective space? A case for appraisal. Cogn Emot 20(1):92–113
28. Stanton NA, Edworthy J (eds) (1999) Human factors in auditory warnings. Ashgate Publishing, Farnham
29. Tractinsky N, Katz A, Ikar D (2000) What is beautiful is usable. Interact Comput 13:127–145
30. Zwicker E, Fastl H (1990) Psychoacoustics facts and models. Springer, Berlin