Towards Informed Metaphor Selection for TUIs

Stefan Oppl  
Kepler University of Linz  
Altenberger Strasse 69  
4040 Linz, Austria  
stefan.oppl@jku.at

Chris Stary  
Kepler University of Linz  
Altenberger Strasse 69  
4040 Linz, Austria  
christian.stary@jku.at

ABSTRACT
In TUI design the selection of metaphors influences user expectations and ease of use. Traditional TUI design processes have not addressed this issue explicitly so far. A reflection of existing approaches for metaphor classification in TUI design helps explaining why TUI metaphors not fitting could mislead users. Based upon these explanations, we could gain empirical insight into negative effects caused by selecting metaphors not fitting the situation of use. The results allow pursuing a metaphor-aware TUI specification process, as they address metaphor selection explicitly, and can be grounded in both, concept development, and empirical findings.

Author Keywords
Tangible User Interface, Metaphor, Specification, Study

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Haptic I/O

ACM General Terms
Design, Human Factors

INTRODUCTION
Interaction based upon metaphors allows users selecting activities appropriate to their context of work and current task [1]. Selection of these metaphors is crucial for the comprehensibility, and thus usability (ease of use, conformance to user expectations, and learnability) of an interactive computer system [2]. In the field of Tangible User Interface (TUI) research, early works (e.g. by Ishii & Ullmer [3]) have already recognized the importance of metaphor. Recent publications [4][5][6] continue to consider this topic relevant.

Several authors [7][8][9] have systematically investigated metaphors for TUI (Tangible User Interface) design. However, they mainly focus on ex-post analytical classifications of TUI metaphors. Approaches targeting towards the specification of design tasks (e.g. the TAC paradigm as proposed by [10]) currently do not consider metaphor selection explicitly.

In this work, we describe the first steps towards a design process that fills this gap. In the first part of the paper, we review and summarize existing work on TUI metaphor classification and identify the design dimension to be considered when selecting metaphors for the elements of a TUI.

In the second part of the paper, we present empirical evidence of the effects of ill-chosen and well-chosen metaphors on TUI interactions. We then describe how the negative effects of ill-chosen metaphors could have been anticipated, once considering the conceptual framework in the course of design. In the final part of the paper we outline how the findings of this paper can be integrated with a structured specification process for TUIs.

METAPHOR SELECTION AS A (T)UI DESIGN TASK
In general, metaphors are pervasive in language and thought. They trigger actions [1] and are bound to human experience [13]. Since shaping also task accomplishment and interaction, they have to be considered by (T)UI designers.

Metaphors have been used to interpret and trigger actions at the computer side [14], and to facilitate human understanding when interacting with computers [15]. Antle et al. [16] show how embodied metaphors facilitate interaction in hybrid interaction spaces in the context of learning. According to their findings, several design goals when using metaphors can be set: (a) facilitate user (work) task accomplishment and (b) reduce cognitive workload of users when interacting with an artifact.

Metaphors emphasize the symbolic significance of particular elements in human(-computer) interaction, e.g., a box representing a relevant set of task items [17]. Even the most concrete and rational aspects of interaction - whether searching, arranging, replacing, or distribution of information - embody meanings, if not social constructions that are crucial for understanding what persons do [18].
Consequently, metaphor-based interaction spaces have to be considered as enacted domains. They are social enactments, as interactions are socially constructed. The persons who bring metaphors to life choose and structure internal and external relationships through a host of interpretive decisions that are extensions of individual properties and style.

Besides the addressed intertwined cognitive and social dimension of metaphor-based design, the virtual and concrete setting of an environment is affected through TUI design:

- Metaphor-based TUIs create and shape activities, hence influencing the way persons accomplish tasks, arrange information, or communicate ideas. Designers have to be aware that embodied metaphors have the capability to shape and guide organized action.

- Using multiple metaphors to organize work tasks and interactions tap different dimensions of a situation. Specific situations might lead to specific metaphors. Otherwise users will find themselves experiencing unanticipated interaction problems that hinder many developments, and lead to endless restructuring, re-inventing, or re-engineering interaction spaces.

Given the social enactment and shaping user behavior, metaphor-based TUI design enriches traditional development processes substantially [19]: Developers refrain from applying their formal authority, function, and role as a kind of protective input that insulates users and themselves from change. In the course of design they need to encounter human experience and develop shared understanding with users and domain experts.

**TUIs AND METAPHORS**

In an extensive review of existing literature on how to systematically describe (i.e. specify or assess) TUIs, we have identified several approaches, which explicitly consider “metaphor” as a dimension of TUI design or assessment. In the following sections, we briefly describe these approaches and point out how they reflect the findings described above.

**Relevant Frameworks**

Many authors have claimed the importance of metaphors when designing TUIs [4][5][6]. However, only few have explicitly addressed how to systematically approach the topic of metaphor-based design, how to classify different types of metaphors, and how these classes become manifest in existing systems.

Underkoffler & Ishii [7] have been the first to systematically approach the mapping between physical and digital world. They propose a continuum of object meanings to classify the elements of a tangible interface. Koleva et al. [8] propose a framework for TUIs in which they use the “degree of coherence” as the primary dimension for classifying TUI elements. The “degree of coherence” describes to which extent “linked physical and digital objects might be perceived as being the same thing”. Finally, Fishkin [9] proposes a taxonomy for TUIs, in which “metaphor” is one of two dimensions. According to Fishkin, “metaphor” describes, whether the “system effect of a user action [is] analogous to the “real”-world effect of similar actions” [9].

None of the authors relate TUI design to their frameworks. They rather use them as means for ex-post classification and analysis. In the following, we examine the frameworks’ potential to be applied in the course of designing a TUI.

**Meanings of Tangible Objects**

Underkoffler & Ishii [7] propose a continuum of object meanings that allows classifying the physical elements of TUIs. They restrict the scope of their classification to be applicable only to elements of “luminous tangible systems”, a class of systems that now commonly is referred to as “interactive surfaces” [5].

In the center of their continuum they define the class “noun” to represent objects directly corresponding to their digital counterpart. Physical manipulations of these objects directly affect the representation of the manipulated property in the digital world. On either side of the continuum, either the physical properties of the object (i.e. what an object is), or the ways it can be manipulated (i.e. what an object does), is getting more and more irrelevant for the mapping to the digital world. Stripping away the manipulation of an object, classification becomes “attribute”, where only parts of the physical properties of the object affect the digital representation. It finally leads to “pure object” where only the “being” of an object is represented in the digital world. On the other side of the continuum the physical properties of an object are not of interest. Classification here starts with “verb”, where altering the digital representation is tied to some manipulation of a physical object, which, however, is not related to the object or phenomenon that is manipulated in the digital representation. Finally, the object becomes a “reconfigurable tool”, where the object’s physical appearance gives no hints at all of how it can be used and the object can be applied to trigger multiple manipulations in the digital world.

**Degree of Coherence**

Koleva et al. [8] classify TUIs along a continuum describing the “coherence” of the TUI elements. The “degree of coherence” describes the strength of coupling physical with digital objects, i.e. to which extent they are perceived “as being the same thing”. Starting with strong coherence the initial classification is the “illusion of same
objects”. Users do not distinguish among physical and digital properties of an object here.

Stepping down the coherence scale the subsequent classifications are “projection” and “proxy”. Both maintain a close coupling of physical objects with digital representations. However, they are still perceived as separate phenomena by the users. “Projections” are used whenever certain properties of a physical object directly map to the digital representation. The latter only exists, once the physical object is present. “Proxies” are still permanently coupled to certain digital representations. They allow for manipulation. The existence of the digital representation, however, is not necessarily bound to the presence of the physical token. “Identifiers” are objects representing only digital information while being physically present. However, they do not allow manipulating this information. “Specialized tools”, in turn, do not have information assigned to permanently, but are used to perform specific manipulations using the digital representation. At the lower end of the coherence-scale “general-purpose tools” do neither represent specific information, nor indicate specific manipulations of the digital representation.

**Taxonomy of Tangible Interfaces**

Fishkin [9] proposes to use a two-dimensional taxonomy when classifying TUIs. The first dimension – “embodiment” – describes how “close” modalities of input are tied to output (with the scale ranging from “distant” of “environment” and “nearby” to “full”, where e.g. “nearby” would be information projected onto a table surface just underneath a token that is used to manipulate or control this information). The second dimension – “metaphor” – describes to which extent user actions cause the same effects within the system as they would cause in the physical world.

A metaphor classification of “full” describes TUI objects not perceived differently from physical world objects in both, their appearance, and the way they can be manipulated. Digital information is integrated in a natural way, i.e. not disrupting the perception when interacting with a object from the physical world. TUI objects use a “Noun and Verb”-metaphor, in case both, the appearance of the TUI object (“noun”), and its usage “verb” resemble a corresponding physical world object and interaction. Users, however, still perceive the TUI and “plain” actual world objects to be in different contexts (i.e. a TUI object is not usable in the actual (“real”) world and vice-versa). A metaphor of “verb” decouples actions from objects and shows analogies between TUI and “plain” actual world only on an interaction level. While the object used in the TUI does not have an actual world counterpart (as would be the case for a “generic” information-container token), the manipulations that can be performed on this object correspond to actual world interactions (i.e. hand over the token to give information to another person). Accordingly, an object with metaphor of “noun” shows analogies regarding its appearance, but cannot be manipulated or used as in the actual world. Finally, the metaphor classification “none” is used for objects having no correspondence in the actual world at all, neither regarding their appearance, nor their usage or manipulation.

**Discussion**

Each of the reviewed approaches aims to classify TUIs along the used metaphors. Although focusing on different aspects, the approaches share some concepts we are going to discuss.

None of the frameworks is applied to entire TUIs, but rather to objects they consist of. Thus, a TUI can be related to multiple classifications. Recognizing that the question arises, whether “consistency” (in terms of realizing metaphors in a way that all tangible objects can be assigned to a single specific class) is an issue here. While this aspect is beyond the scope of this paper, further research is required to examine the use of different metaphors within a specific TUI-based system.

The identified continua follow a similar approach when looking at the classes of metaphors. Figure 1 presents a mapping between the classifications found in the authors’ descriptions. Fishkin [9] has provided a similar mapping between his approach and the Object Meanings of Underkoffler & Ishii. Our mapping deviates from his understanding in one detail:

We do not agree to Fishkin’s mapping of Underkoffler & Ishii’s “noun” class to the “noun”-class of his taxonomy. According to Underkoffler & Ishii, “noun” is a “maximally ‘real-world’ object reading”, where objects are „fully literal, in the sense that they work in their luminous-tangible context very much the way objects ‘operate’ in the real world“. This understanding would indicate a mapping to Fishkin’s “Noun+Verb” or even “Full”-Class. We consider Fishkin’s work inconsistent to that respect. He claims the building tokens of “Urp” to be an example for “Noun+Verb” in his taxonomy, where the main elements in “Urp” are mentioned as examples for “Noun”-meanings in Underkoffler & Ishii’s work.

The mapping presented in Figure 1 points out the commonalities and differences in the three described classifications:

On the end of the continuum representing “stronger” metaphors, all approaches define a class of TUI objects that cannot be distinguished from actual world objects with respect to both, their appearance, and usage - digital information is integrated with the physical objects. On the other end of the continuum, representing “weaker” metaphors, all approaches describe TUI objects that do not resemble “real” objects, neither in appearance nor in usage.
Within the continuums, the authors differ with regard to focus and granularity. Koleva et al. focus on “stronger metaphors” and define a more fine grain classification here. In turn, Underkoffler & Ishii provide more detail on “weaker” metaphors. 

The study is based on a system that has been developed in the context of supporting negotiation of meaning and development of a common understanding for groups by using concept maps [11]. The system is implemented using a tabletop interface with tangible building blocks that represent the concepts. Associations among the concepts are projected onto the table surface and can be set and removed by specific user interactions.

As part of a more extensive empirical evaluation, the interactions necessary for the removal of associations in a concept map have been examined with respect to the usage of different metaphors. A physical eraser represents the token used for removal. Hence, the appearance metaphor has been fixed. Two different forms of interaction have been implemented utilizing the eraser token. They differ in their usage metaphor.

In the first case presented here the eraser token was designed to be used to switch from the current to the “erase mode”, by placing it somewhere on the surface. After the eraser had been put onto the table surface, the system switched to erase mode, and additional interaction was required to actually remove associations - no explicit usage dimension in the metaphor, functionality is tied to plain “being” [7] of the eraser.

Hence, the eraser tools have been built upon a purely appearance-based metaphor, ignoring the dimension of usage. Consequently, the eraser (physical appearance) switched to the connection-removal mode, but could not be used to actually erase connections (indicated interactive abilities).

In the second case, the erase interaction was altered to become stateless. It direct led to removal when placing the eraser on specific associations (usage metaphor: “erasing”). The interaction with the eraser token thus matched the metaphor suggested by its physical appearance (i.e. connections actually were erased). In Fishkin's metaphor the new version of the tool is classified as “Noun+Verb”.

The following hypothesis was tested for both cases: The metaphor the eraser token is based upon allows users in an intelligible way removing connections (H1). H1 had to be rejected, in case users were unable to remove associations just by applying their assumptions about how the eraser token has to be applied.

**Study Design & Methodology**

Overall, 128 people (102 male, 25 female) participated in 57 mapping sessions (in groups of 2-3 people each). The participants were students of a curriculum in business information systems and had no prior experience in using the system. We recorded all sessions using video. The videos were the basis for transcribing problematic interactions between users or with the system (following the interaction analysis approach described by [18]).
Measures
To validate or falsify H1 two quantitative measures have been used alongside with qualitative findings from interaction analysis and user feedback in the questionnaires.

Misinterpretations (M) – measured by number of times the eraser has not been used in the way it was intended to be operated by the designer. The measure is represented as percentage of total eraser usage in the respective session. High values reveal various misinterpretations, and thus indicate problems applying the metaphor.

Number of times the eraser has been used for association removal in favor to other means of removal (ER) – measured by number of times the eraser has been used to remove an association in favor to alternative ways of removal (like “undo”). Alternative ways cause more interaction effort (take longer to be performed and/or have side effects on the remaining concept map). The measure is represented as percentage of total amount of removed associations. High values show intense usage of the eraser, and indicate an comprehensible metaphor.

Case 1: Noun-based metaphor
The first version of the eraser feature was tested in 36 mapping sessions with 91 participants (73 male, 18 female).

Across all sessions a total of 123 (successful and unsuccessful) removal activities occurred. In 19 cases users applied the eraser token, in the remainder of 104 cases users performed the removal using more complex ways. Thus, measure ER is 15,4%.

In 17 of the 19 situations in which the eraser token was used, connection removal was unsuccessful due to misconceptions of how to use the token. Thus, measure M is 89,5%. Video recordings of these situations show how users acted confused and felt unsure on how to proceed after selection. Additionally, 60 of the 91 participants explicitly mentioned the eraser token as the most confusing feature when asked in an open question which aspect of the system they found least intuitive. Based upon these results, H1 has to be rejected for case 1.

Case 2: Noun+Verb-based Metaphor
37 participants (30 male, 7 female) tested the redesigned version the removal tool in 21 sessions (with 6 participants taking part in more than one session).

Across all sessions users performed a total of 130 (successful and unsuccessful) removal activities. In 103 cases the eraser token was used; the remainder of 27 cases was performed using one of the alternative ways for connection removal. Measure ER is 79,2%. ER is significantly higher for case 2 (usage of eraser codified with ‘1’ and alternative ways with ‘0’, one-sided Wilcoxon-test for unpaired samples: \( W = 2895,5, \) \( p < 0,005 \)).

In none of the 103 situations the eraser token was used, connection removal was unsuccessful due to misconceptions of how to use the token. Hence, measure M is 0%. M is significantly lower for case 2 (misconception codified with ‘1’ and no misconception with ‘0’, one-sided Wilcoxon-test for unpaired samples: \( W = 103, \) \( p < 0,005 \)).

In the written feedback following each session none of the participants mentioned the association removal feature or the eraser token feature, although being asked which aspect of the system they found least intuitive. Based upon these results, H1 can be accepted.

Discussion
The metaphor that has been used to implement the association removal feature was the eraser one. In the first case, however, this metaphor was only applied to the token itself, but not to the corresponding interaction. The eraser had led to state changes of the system, but had no immediately visible effects on the object to be manipulated (= the association to be removed). The intended form of interaction has not been transparent to the users of the system. They misinterpreted how to use the token or refused to use it at all (even after initial training of how to use the eraser token correctly).

In the second case, the eraser metaphor was applied to both, the token and the according interaction. The removal feature was triggered by the natural form of interaction with an eraser, namely rubbing it on the part of the surface where something should be removed. This version was used significantly more often, and also perceived significantly better than the first version.

Following the mapping between Koleva et al.’s and Fishkin’s classification, these findings could have been anticipated already during design. “Noun”-based metaphors only should be used for “identifiers”, not for “tools” (as it is the case for the first version of the eraser token). The implementation was changed in the second version to fit the “verb-based”-metaphor anticipated by users, thus making the token act as a tool. By considering “verb” and “noun” aspects of the chosen metaphor, higher coherence [8] could be reached.

The eraser token uses a strong appearance-based metaphor (using a physical eraser as a TUI object) that imposes certain affordances [27] (using the object as a tool, like a physical eraser, i.e. to rub on the drawings to be erased), which were not met by the noun-based implementation of the removal tool. The users intuitively extended the appearance-based metaphor with a usage-based metaphor. Hence, they upgraded the TUI object to a more “coherent” classification than it actually provided.

Consequently, it seems to of importance not only to select a suitable metaphor for a feature to be implemented using a TUI. It is necessary to consistently apply the
metaphor to both, the physical object embodying the tangible token, and the usage of the token effecting the digital representation.

CONCLUSIONS
Using a mapping between three existing classification schemes we have examined the role of metaphors in the design space for TUIs. Based on an empirical study we could illustrate the problems arising once metaphors are not well reflected for TUI design.

The classifications of Underkoffler and Ishii [7], Koleva et al. [8] and Fishkin [9] and their mapping provide starting points of how metaphors can be selected during TUI design. The distinction between appearance-based and usage-based metaphors appears to be a suitable approach to explicitly consider metaphor selection during TUI specification using the TAC-scheme [10]. The TAC-scheme enforces the definition of both, the (physical and digital) representation of a TUI element and its behavior. Depending on the intended purpose of a TUI element [8][9], such as the “removal tool” in the case of the eraser token, specification might either focus on metaphor selection for behavior (usage) or representation (appearance), while still considering the other dimension to ensure the metaphor’s intelligibility.

In future work we will focus on the development of a metaphor-aware TUI specification process based upon the reviewed metaphor classification approaches and the TAC-scheme. The proposed specification process needs further empirical work for justification in “real” world scenarios.

REFERENCES
15. Pirhonen, A., Brewster, St., Holguin, Ch. Gestural and audio metaphors as a means of control for mobile devices, Proceedings CHI 02, ACM 2002.