

Affordances in Conversational Interactions with Multimodal QA Systems

Andreea Niculescu

HMI Group, University of Twente
niculescuai@ewi.utwente.nl

Abstract. Implementation of adequate conversational structures is a key issue in developing successful interactive user interfaces. A way of testing the adequacy of the structures is to prove the correct orientation of each communicative action towards a preceding action. We refer to this orientation leading to a certain response as the affordance of the communicative action. In this paper we present a case study where affordances of implemented conversational structures (including verbal and graphical elements) in a multimodal medical QA system are identified applying Conversation Analysis (CA) tools and tested using the Cognitive Walkthrough (CW) method. The CW method was modified to fit the conversational approach and tested with five expert evaluators. Results showed that the affordance analysis helps detecting inefficient constructions leading to disruptions in the dialog flow, spots unnecessary functions and provides important insights on systems easy-of-use.

Keywords: Affordances, Conversational Interactions, Multimodal QA systems, Usability Inspection, Cognitive Walkthrough.

1 Introduction

This paper discusses the design evaluation of a multimodal QA system from the perspective of the affordance concept.

We claim that many problems associated with natural language based interactions are originated from the lack of deeper understanding of underlying conversational structures. Typical Graphical User Interfaces (GUI) use preponderant visual interaction elements. Therefore, the affordance design focuses mostly on visual elements, regarding verbal units as simple cognitive support for the graphics.

Multimodal QA systems are a special case of GUIs in which graphical elements are combined with large text units. The rapport between verbal and graphical elements is here reversed: the text units are the one supported by graphical elements being the quintessence of the QA interaction. Consequently, we believe that interaction designers should include more extensive analyses of verbal affordances when designing interactive systems based on conversational structures, like QA or dialog systems.

Since the concept of affordance has been often subject of intense controversies in HCI debates once introduced by Donald A. Norman in [1], many researchers

such as Gaver, Hartson and Norman himself struggled with several definitions and categorizations in an effort to clarify the concept making it operational for evaluations. Section 2 comments in details these theoretical considerations. In section 3 some short examples of practical applications of affordances are presented followed by section 4, where the affordance concept is integrated in the framework of the conversational analysis protocols. The case study, including methodology, short system overview, questionnaire and scenario design are presented in section 5. The results are largely discussed in section 6. This paper ends with conclusions containing the result summarization and improvement suggestions for the QA interface.

2 What Are Affordances?

The concept of affordance was developed by the American psychologist J.J. Gibson in [2] and [3] it is a significant part of his ecological theory of direct perception. Gibson defined the term "affordance" as latent action possibilities existing in an environment independent of the individual's ability to perceive them. These action possibilities are in relation with the actor's capabilities of action being independent of his culture, prior knowledge or expectations. Any substance, any surface, any layout has some affordances with respect to a certain actor. According to his theory, the action possibilities indicated by affordances are perceived visually in a direct way that does not require mental information-processing activity, i.e. the immediate perception of the environment will inevitably lead to a certain action.

Gibson's theory of non-conscious information pick-up was criticized as it fails to explain how actors assign meaning to what they see deciding whether to perform an action. Even if the existence of affordances is independent of the actor's experience and culture, the ability to perceive such affordances may be dependent on them. Therefore, the actor may need to learn first to discriminate the information he gets from the environment in order to perceive it directly [4]. Although Gibson rejected the involvement of mental activities in the process of direct perception, he conceded that even for the most "basic" affordance perception might need to somehow develop - a clear suggestion that learning could be involved.

Other criticisms to Gibson's theory refer to the fact that affordances are defined as being relational but the nature of the relation between actors and environment are not further discussed. Even though the theory presents some illustrative examples of affordances, it doesn't provide an analytic way of identifying affordances[5].

Despite all the criticism, Gibson's theory of affordance brought radical changes in the field of perceptual psychology and was successfully adopted as a key concept in other fields, such as cognitive science, robotics, artificial intelligence, design etc.

In the HCI field the term "affordance" was introduced by David A. Norman in his book "The Psychology of Everyday Things" [1]. Norman adopted Gibson's

concept and used it to address design aspects of artifacts considering technology as part of the environment. He defined affordances as physical, "perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used. [...] Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. [...] When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction needed" [1].

However, even though Norman borrowed the term from Gibson he disagreed with his theory about whether the mind processed or simply "picked up" information (see [1], [6] and [7]). Consequently, Norman departs from Gibson's theory and considers affordances perceived properties of objects that may or may not exist. The perception is determined not by the action capabilities of the actor as Gibson says, but by his mental and perceptual capabilities. Moreover, past knowledge and experience are tidily coupled with affordances in Norman's view [4].

For Norman, the notion of affordance becomes a mixture between actual (shape, material, color) and perceived properties of the object where the perceived properties are in fact the suggestion of how the object might be used.

The lack of separation between physical properties and perceptual information about their use created a rather ambiguous definition of affordance and generated large discussions about the meaning of the term in the HCI community.

A substantial contribution to clarify the concept bringing the Gibsonian thinking back in the HCI world was made by the interaction designer William W. Gaver. Similar to Gibson, Gaver argued that affordances do exist independently of their perception, being ontologically but not epistemologically relevant. He defined the affordance concept as the property " [...] of the environment relevant for action systems" [8] and proposed a taxonomy where the affordance concept is separated from the perceptual information available about it.

In Gaver's framework affordances (aff.) and their perceptual information (per.inf.) are defined as entities taking binary values such as "*yes*" and "*no*". Their combinations result in four types of affordances called: perceptible, hidden, false and correct rejection.

- a. Perceptible affordances: (aff.: *yes*, per.inf.: *yes*) offer a link between perception and action by signaling a possible action in a visible way.
- b. Hidden affordances: (aff.: *yes*, per.inf.: *no*) offer no link between perception and action; actions are possible but there is no signal acknowledging their existence (e.g. a hidden door).
- c. False affordances: (aff.: *no*, per.inf.: *yes*) offer a link between perception and a non-existing action possibility; actions are mistakenly signaled as being possible (e.g. a door might appear to afford opening, but it won't afford if it's locked).
- d. Correct rejection: (aff.: *no*, per.inf.: *yes*) refers to the situation where no action is afforded or signaled (i.e. no affordance).

While conveying the categorization to interaction design field designers should avoid false and hidden affordances as being a sign of weak design: false

affordances bring users on a wrong path while hidden affordances waste resources, as users will probably encounter difficulties on detecting their existence. Instead, designers should concentrate on making affordances perceptible or creating situations where the lack of affordance is correctly rejected.

Another attempt to extend and refine Norman's concept of real and perceived affordance came from Hartson [9]. Hartson proposed four complementary types of affordance in the context of interaction design and evaluation: cognitive, physical, sensory and functional affordance.

Cognitive affordance -corresponding to Norman's perceived affordance- is associated with the semantics of the interfaces and refers to design features that help users knowing something (e.g. the label of a button indicating what will happen if a user clicks on it).

Physical affordance -corresponding to Norman's real affordance- is associated with characteristics concerning the "operability" of the interface, and refers to design features that help users to accomplish accurately a physical action in the interface (e.g. the size of a button that is large enough to allow users to click on it).

Sensory affordance is related to "sense-ability" characteristics of the interface and targets design features that help users perceiving (e.g. seeing, hearing, feeling) something (e.g. the font size of a label). Sensory affordance plays a critical supporting role to cognitive and physical affordances.

The last category, functional affordance addresses design features that help users to accomplish work (e.g. the internal system ability to sort numbers invoked by a user who clicked the 'sort' button [9]).

There are several other interesting interpretations and formulations of the affordance concept but due to obvious space limitation we discussed in details only a few of them that we considered as being the most relevant to our analysis.

3 Practical Dimensions of Affordances

The study of affordance goes beyond theoretical speculation; authors like Vainio et al. [10] validated the affordance concept as part of an interesting empirical study. They showed that participants during several tests could identify objects faster if they were congruent with an observed action prime (e.g. power grasp - power grasp compatible object) rather than incongruent (e.g. power grasp - precision grasp). They concluded that motor knowledge plays an important role in object identification and, consequently, action-related information associated with an (graspable) object is an inseparable element of that object's representation.

In the HCI field the affordance concept has found its practical settings as design model and analysis tool for physical and graphical user interfaces.

Sheridan & Kortuem [11] proposed an affordance-based design model of physical interfaces for ubiquitous environment. They proposed an experimental method to study object affordances showing how the method can be applied to the design of concrete physical interface artifacts.

Luo et al. [12] used the affordance design model as theoretical basis and methodological underpinning to evaluate an e-learning program on mammogram reading.

Hartson [9] explored the relationship between the affordance types associated with usability problems and provided examples and a methodology scheme for practitioners on how to identify affordances issues involved in flawed design cases.

However, the theory of affordance and its practical applications concern mainly the visual perception of environment objects analyzing verbal element only marginally, e.g. only when they are meant to support visual elements (see Hartson's cognitive affordance). Since new media technologies such as interactive information systems like QA or dialog systems are design artifacts that use elaborated conversational structures with preponderate verbal text elements there is a strong need to consider such elements as integral parts of the interaction design. Therefore, we propose a design evaluation that uses the affordance concept to analyze text units and graphical elements.

4 Affordances in Conversational Interactions

An important characteristic of conversational interactions in general is the fact that they are deeply anchored in the cultural context of use; that means they are based on conventions and constrains of the socio-cultural environment following a rigorous protocol course.

Contrary to Norman who argued that the socio-cultural world is placed outside the domain of affordance [6], Gaver emphasized the role of the culture together with other factors such as experience and learning involved in the process of perception of affordances [8]. We also believe these factors are not to be considered affordances but have an important function in making affordances visible.

A way of detecting these factors and implicitly affordances in conversational interactions is to apply conversational analysis (CA). In CA, conversations can be considered "environments" where certain types of actions such as gestures, mimics, verbal statements are afforded in certain circumstances - in terms of non-violating coherence principles and cultural constrains. Interlocutors can express their communicative intentions both verbally (through speech) and non-verbally (through gesture and mimic). Analyzing the organization of each conversational sequence it can be determined what kind of action possibilities (affordances) have the participants in a certain moment of the conversation. We consider these affordances from a pragmatic perspective, i.e. action possibilities oriented to achieve a certain goal.

Conversations are by nature interactive and follow a relatively strict turn-based protocol. We address in this paper only the case of closed-domain question-answer interactions since the system analyzed in our case deals with this type of interaction.

In general, question-answer conversations are fully structured in adjacent pairs, meaning that all exchanged turns are functionally related to each other in such manner that the first turn requires a certain type (or range of types) of

second turn [13]. The adjacent pairs are grouped in three separate categories, corresponding to a conversation initialization, termination (both containing greeting-greeting pairs) and body sequence (containing question-answer pairs). These three categories form together what we call conversation protocol.

4.1 Conversation Initialization and Termination

Starting and ending a conversation are levels of phatic communication with social functions: they are responsible for establishing rapport or quitting the interaction "circle" in a polite way.

A conversation usually starts with a signal showing the readiness to engage in a conversation. Such signals are salutation forms, self-presentation (if the interlocutors haven't met before), non-verbal gestures (hand shake, hugging, kissing, hand waving, gazing), mimic (smiling), changing the corporal position towards the interlocutor, etc.

A similar protocol for ending the conversation includes farewells and thanks exchanging, waving arms, hugs, kisses, glancing away, re-orienting body posture away from interlocutor etc.

Each performed action affords in principle a similar one in return: a greeting, a smile, a self-introduction affords symmetrical responses from the interlocutor. However, the realization of conversation initialization might differ across culture taking into account participants' gender, age, social position and degree of acquaintance. For example, in Western cultures a stretched out hand will afford hand-shaking; in Muslim countries such a gesture will afford hand-shaking usually if both interlocutors are men; women instead will press their hand upon the chest to signalize salutation response avoiding at the same time the direct contact with the opposite gender.

4.2 Conversation Body

The conversation body contains the essential part of the interaction, namely the information exchange (also called informational communication).

The information exchange may start with a short explanation of the intended nature of the conversation-to-be preceding (the first question-answer exchange). At this point a common ground (a set of propositions that make up the contextual background for the utterances to follow) can be established.

From a pragmatic point of view a question may afford following responses: a matching answer, acknowledgment of ignorance, suggestion for asking someone else (re-routing), intermediary questions to clarify a previous question, postponement, refusal to provide an answer, feedback showing that the question was understood or a request for time to process the question, etc.

In case of miscommunication repair strategies occur in form of explanatory adjacent question-answer pairs.

The turns can be accompanied by non-verbal cues such as gestures and mimic used to emphasize the content; e.g. gaze signalizes attention and readiness for interaction, rising eye-brow shows surprise, smile acknowledges agreement, etc.

Question and answer pairs must respect the coherence principle by being semantically and meaningfully related to each other. In order to achieve coherent information exchange, syntactical features such as anaphoric, cataphoric and deictic elements may be used. Also logical tense structure, as well as presuppositions and implications connected to general world knowledge are deployed to coherently connect answers to questions[14].

In common practice interlocutors do not perform their utterances at the same time. Speakers usually take turns to talk. Overlapping and simultaneous talk is generally seen in Western cultures as unpleasant. The turn-taking usually occurs at the utterance ends, often signaled by silence.

Interruption might be allowed if one of the interlocutors signaled verbally or by gesture the wish to take the turn[15].

5 Method

Even though conversational interactions with multimodal systems differ in many aspects from their human counterpart, they generally follow the same conversation protocol consisting of initialization, body sequence and termination. This similarity is intentionally simulated by designers in order to increase the system's easy-of-use and to make users' answers predictable.

As theoretical framework we adopted Gaver's taxonomy to identify affordance values and Hartson's scheme to establish affordance types. The analysis followed the conversational protocol steps described above.

The test was carried out using the Cognitive Walkthrough (CW) method. The method is a usability engineering tool meant to help designer teams to quickly evaluate interaction systems from early stages of development. It does not require a fully functioning prototype (as it is the case with the system on which the test was performed) nor users involvement. CW emphasizes cognitive aspects, such as learnability by analyzing users' mental processes required for each step[16]. Design experts perform the test taking into account the potential user perspective with the purpose of identifying problems that might arise during the interaction. After each scenario the experts are asked to answer questions stated in a questionnaire.

5.1 The IMIX System

Our study was performed on IMIX, a multimodal interactive QA system for medical queries with extended follow-up questions functionality.

IMIX was developed for educational purposes and can deal with medical encyclopedic questions, i.e. general questions that do not require expert knowledge like diagnostic questions or complex medical analysis [17]. The system's users are expected to be people with no professional knowledge of the medical domain. They will probably make use of such services only occasionally. No special training is required to interact with IMIX.

The system is primary text-based but allows users to use optionally speech¹. Attached to IMIX is a talking head called Ruth.

¹ During the test only the text-based modality was deployed.

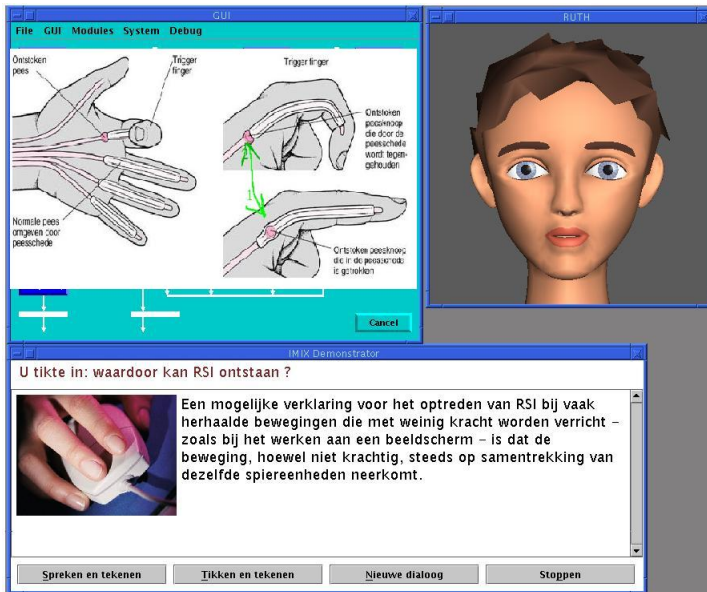


Fig. 1. Screenshot of the IMIX system

Cognitive research into multimedia has shown that the use of text combined with pictures may substantially contribute to the user's learning process of the presented material [18],[19]. Therefore, the IMIX answers contain text combined with suitable static images. The answers are made-up by matching the query to document fragments from the data base. The text answers may be both spoken by Ruth or displayed on the screen. Optionally, follow-up questions can be formulated as text, speech or drawing [20].

5.2 Prior Issues to the Test

Before starting the test the evaluators got paper sheets containing preliminary information about the test goal, a short description of the term "affordance", a detailed explanation of the human conversation protocol and a general description of the IMIX system.

Afterwards, the evaluators received the scenarios containing specific tasks to accomplish, a list of correct actions required to complete each of these tasks and a separate questionnaire for each scenario.

5.3 Scenarios and User Profiles

Three scenarios covering the conversational structures implemented in the IMIX system were developed. We tried to design the scenarios as pleasant and humorous as possible in order to achieve an enjoyable interaction. Each scenario focuses on a specific task.

In the first scenario evaluators were asked to put a single question and analyze the corresponding conversation protocol. The scenario identifies the situation of a naive user with little medical expertise who uses IMIX to find out what means Repetitive Strain Injury (RSI).

In the second scenario the evaluators had to concentrate on the special case of follow-up questions using drawing and/or typing options. The user profile of this scenario corresponds to a subject with search engine expertise and interests in the medical domain. He uses IMIX to find information related to liver functions.

The third scenario addresses repair and meta-communication strategies when the answer to a question is not found. The user profile addresses an expert user who uses IMIX for entertaining purposes. He is seeking for information about the SARS virus.

The evaluators were asked to keep in mind the aim of the test: to look at the way the user is invited to interact with the system and NOT at the answer quality he/she might get back. They also had the possibility to repeat a scenario several times if they wish so.

5.4 Questionnaire Design

For each scenario a separate questionnaire was developed. The questionnaires were designed in accordance to Cognitive Walkthrough (CW) method[21]. However, the questions were adapted to fit the special case of verbal interactions and grouped in three units, each one corresponding to a separate conversation protocol category.

The purpose of the questionnaires is to detect affordances of the conversational sequences implemented in the protocol. The structure of the questions is similar for each unit. Evaluators have first to detect elements signaling the current protocol. Then they have to anticipate users' re-action given a certain conversational sequence. Furthermore the evaluators have to determine whether the users' responses are acknowledged by the system and how they will perceive this feedback. Eventually each question unit ends with a question about potential violations of the conversation protocol. This last question is meant to catch issues that might have "escaped" the evaluator's observation.

5.5 Pilot Study

Before starting the experimental run a first pilot study with one expert evaluator was accomplished. From the pilot study three main observations could be gathered:

- 1) The relatively high difficulty degree of the question demanded the presence of experienced evaluators having some affinity with the CW method.
- 2) Typical CW questions like "Will the user notice the conversational starting signals(as the correct action available)?" are too general. Precise formulations similar to "Will the user understand this signal as an invitation to start a conversation?" seemed to be more appropriated even if the questionnaire size increases, e.g. for each signal one separate question.

Table 1. Classical CW versus Adapted CW

Classical CW	Adapted CW
1. Will the users be trying to produce whatever effect the action has?	1.0 What elements are used to signalize [a certain action]? 1.1 Does the word [...] suggest [a certain action]? 1.2 What kind of statements affords the question [...]?
2. Will users be able to notice that the correct action is available?	2.0 Will the users see there is [a certain] option ? 2.1 Will the users understand this signal as an invitation to [do a certain action]?
3. Once users find the correct action will they know that it is the right one?	3.0 Will the users know how to use [a certain] option?
4. After the action is taken, will users understand the feedback they get?	4.0 Will there be a feed-back to acknowledge the action performed by the users? 4.1 Will the users understand these kind of feed-back? 5.0 Does the conversational protocol get in any way violated?

3) Since the answers to the CW questions are often not straightforward, requiring some deliberation time it seemed wise to record the testing session. In this way a considerable amount of time could be saved and no observation could get lost.

6 Results

The test was completed by five evaluators with design expertise recruited from our department. All evaluators except one were novice using the system. The results of their evaluation are summarized below following the conversational protocol categorization:

6.1 Conversation Initialization

The conversation initialization implemented in IMIX doesn't afford symmetrical response. The conversation's start is signalized by a textual welcome message, a short system presentation, a 'start' button and a talking head emerging from the background gazing and rising eyebrows. At this point the only afforded action is the pressing of the 'start' button to begin the "conversation". No other actions like greetings or salutation gestures in return are afforded, even though according to the conversation protocol a greeting affords another greeting. The occurrence of a signal (greeting message) combined with the lack of an adequate response (no greeting in return) indicates the presence of a false cognitive affordance.

Most of the evaluators agreed that this stage of the interaction has less resemblance to what is normally called a "conversation". The talking head appearance is not a very convincing invitation to talk even if its blinking eyes indicate a waiting behavior. One of the evaluator argued that the presence of speech, e.g. a welcome message read by the talking head would increase the users' feeling of being involved in a conversation.

An adjustment in the head mimic would be beneficial as well: a gazing behaviour combined with smiling is a more appropriate way to start a conversation.

The short system presentation was criticized as being too technical: especially less experienced users would have difficulties in understanding the meaning of having a "multimodal dialog" with the system.

Table 2. Affordances in conversation initialization

Affordance type	Affordance value	Action
Physical & Cognitive	Perceptible	Press 'start' button
Sensory & Physical	False	Click on highlighted words
Cognitive	False	Give symmetrical response

Also the text color should be uniform -one criticism addressed the presence of colored words in the text message, a fact that could mislead internet experienced users to click on it, as it is common on websites with embedded links. This would be a false sensory affordance.

6.2 Conversation Body

The conversation body includes single question-answer sequences, follow-up questions, meta-communication and repair strategies.

Single Question-Answer Sequences

The conversation body begins once the 'start' button is pressed. The users arrive on a new screen where they receive some brief instructions on how to interact with the system.

At this stage of the conversation the users have to choose between two input options: speech or typing. All evaluators agreed that input selection modalities seem to be afforded in a proper manner: the buttons are intuitively labeled and it was estimated that all user categories won't have difficulties to select an input option. A suggestion was made to use additionally explanatory icons like a pen for the typing option and a microphone for the speech option.

It was criticized the presence at the same level of two other buttons: one for the stop option and the other one for the new dialog. The 'stop' button should be placed in a corner -in order to be congruent with the typical design of closing buttons while the 'new dialog' button should be removed as being at the moment functionless and indicating a false physical affordance.

Another remark was to adapt the head position towards the typing input field while users are typing a question in order to increase the interactive feeling and to give a certain feedback. Due to the lack of adequate mimic reactions and synchronization with the current conversation stage the talking head gives the impression that it doesn't belong to the system.

By selecting the option "typing" an input field appears on the same window. The input field affords sentence-like questions as well as keywords. Most of the evaluators -excepting one- considered that the full sentence capability won't be easily perceived by more experienced users; they would probably associate the system's functionality with the one of a typical search engine and consequently would use keywords. The presence of a relatively extended input field is not a clear indication of the expected input and if sentence-like input is desired a short how-to-ask example should be provided. It can be concluded that the input field has hidden physical affordances.

The input field is introduced by the question "What would you like to ask or say?". The designer's intention was to let users know the system is able to

Table 3. Affordances in single question-answer sequences

Affordance type	Affordance value	Action
Cognitive	Perceptible	Chose input options
Physical	Perceptible	Type in the input field
Physical & Cognitive	False	Click 'new dialog' button (first dialog screen)
Cognitive	False	Perform whatever question
Physical	Hidden	Put sentence-like question
Cognitive	Hidden	Perform greeting
Cognitive	Hidden	Perform transition statement
Physical & Cognitive	Hidden	Press 'new dialog' button to type
Physical & Cognitive	Hidden	Press 'follow-up question' button to type

handle different types of statements like full-sentences questions, greetings or even transition formulations ("ok" or "thank you"). But being rather too open, the question suggests it can deal with any kind of statements which certainly is a false cognitive affordance.

On the other hand, most of the evaluators -excepting one- concluded that ironically, especially experienced users would not be aware of what exactly they can utter, e.g. greetings and transition statements affordances remain hidden, as nothing clearly indicates their possible usages.

We continue the analysis considering the case where a naive user will enter a greeting. The system will logically respond repeating the same question ("Hello! What would you like to ask or say?"), but won't indicate how to continue the dialog as the input field disappears. So far a direct answer is not afforded. The users need to press the button for either new-dialog or follow-up question in order to get to the input field to type in, a fact that complicates the conversational flow.

None of the labeled buttons suits semantically the actual conversational situation -the 'new dialog' button should be used in situations where a dialog session re-initialization is wanted, while the 'follow-up' button refers to situations where users are looking for more detailed information in the medical answer. Therefore, it can be concluded that both buttons afford in this conversational sequence a hidden action, namely to allow users to get back to the typing field.

Follow-up Questions

After receiving the first answer the users have the option to continue the information exchange on the same topic by selecting a 'follow-up question' button. The button is labeled with a text indicating that users can type or point at something in the answer. However, the pointing option is not intuitive and has not a specific usage indication. Besides, not only pointing but also drawing is supported, a fact that the label doesn't specify. All evaluators agreed on the fact that all user categories wouldn't know what the option does and how to use it. Moreover, it is not clear which advantages it has compared to the typing option. Therefore, we identify here a hidden physical affordance.

It is also not very clear the way a "drawn" follow-up question is entered for further processing. Since the 'ok' button located in the proximity of the input field can be also used for this purpose, the evaluators concluded the button has hidden physical and cognitive affordances.

Table 4. Affordances in follow-up questions

Affordance type	Affordance value	Action
Physical	Perceptible	Type in the input field
Physical	Hidden	Use the mouse to "drawn" a question
Physical & Cognitive	Hidden	Use the 'ok' button to enter a "drawn" question

There is a feedback to acknowledge the waiting pause and the users' query but the feedback is not specially meant for follow-up questions, fact that should not disturb the communication flow.

Meta-Communication and Repair Strategies

When the answer of a question is not found the system displays a message requesting rephrasing. The function of the rephrasing request should additionally help users to become more successfully in finding the desired information.

All evaluators found the request not supportive at all; according to their estimation even expert users would experience problems rephrasing their question.

After the rephrasing request users can choose between the follow-up question (in the form of typing or pointing) or the new dialog option in order to get cumbersome back to the typing field. Both options were considered inadequate for this particular stage of the conversation. Just like in the follow-up paragraph these two buttons indicate the presence of hidden physical and cognitive affordances.

Table 5. Affordances in meta-communication

Affordance type	Affordance value	Action
Cognitive	Perceptible	Rephrase question
Physical & Cognitive	Hidden	Press 'new dialog' button to rephrase
Physical & Cognitive	Hidden	'Follow-up question' button to rephrase

Conversation Termination

The interaction can be interrupted by clicking the 'stop' button. A real conversational termination is not afforded. Users don't have the possibility to verbally express the intention to leave the conversation, as no typing field was designed at this stage of the conversation. They could click on the 'new dialog' button and type a farewell greeting as the system affords such statements. But this option seems rather a less logical as nobody will probably think to start a new dialog when in fact he/she wishes to stop it. Besides, even if the system replies logically to the farewell greeting it doesn't allow a verbal termination of the conversation.

There is no feedback to acknowledge the end of the conversation and users get the general impression of a system crash by clicking the 'stop' button. We certainly face a conversation protocol violation.

Table 6. Affordances in conversation termination

Affordance type	Affordance value	Action
Physical & Cognitive	Perceptible	Press the 'stop' button
Cognitive	False	No symmetrical response

7 Conclusions

Extrapolating the affordance definition given by Gaver we considered in this paper interactive information systems as artificial environments where verbal and graphical elements are artifacts leading users to perform certain actions. Therefore we proposed a design evaluation in which not only graphical but also verbal elements can be analyzed under the framework of the affordance concept.

The results of our experiment revealed several inefficient structures that could be identified analyzing affordances of conversational structures. The conversation initialization and termination implemented in IMIX do not afford a symmetrical response from the user, perturbing the natural dialog flow. Systems question formulations are too open, a fact that might generate false expectations or disorientation. The labeling of buttons should reflect the actions induced by the buttons. A question should automatically generate a response environment avoiding unnecessary pressing of additional buttons. The system reactions should be consistent at all appearance level, i.e. verbal, mimic, gestures.

The study of affordances also showed unnecessary functionalities that might be removed or adapted in order to become useful. For example, the presence of buttons leading to certain actions should be in accordance with the conversational sequence they are designed for: it makes no sense to start a "new dialog" when no other dialog had been started before. The affordance of certain conversational structures like greetings in the middle of an interaction shows a cooperative behavior. However, it is unlikely that someone would use greetings at that particular conversational stage. Special features like pointing or drawing on a virtual surface should be briefly introduced to users. It is rather unexpected that someone should use an unfamiliar option to ask questions when he/she has more natural choices like typing or speaking.

The affordance analysis also provided important observations about the system's easy-of-use. Users may not understand the system's description, as it seems to be too technical, may not be aware of its full sentence capabilities, may not know whether other transition statements are allowed, may experience difficulties using the pointing/drawing option or rephrasing their questions and may probably feel annoyed when they expect to be able to type a question and no input field is provided.

Last but not least the affordance analysis of verbal elements proved to be beneficial and confirmed our initial claim that many problems associated with natural language based interaction are originated from the lack of deeper understanding of communicative structure: most of false and hidden affordances identified were cognitive nature (6 pure cognitive and 6 physical-cognitive out of a total of 14).

We concluded that understanding affordance of verbal and graphical elements and being aware of their roles in conversational interaction design can help practitioners in diagnosing usability problems from early stage of development, since the affordance analysis using CW methods provides a useful and informational rich perspective for qualitative evaluations of prototypes without implying costly user studies.

Acknowledgments

The author is grateful to Boris van Schooten and Rieks op den Akker for interesting hints and discussions regarding this research, to Betsy van Dijk for useful comments during the pilot study, to Ronald Poppe, Olga Kulyk, Bart van Straalen and Dennis Reidsma for participating in this experiment and to Hendri Hondorp and Dennis Hops for helping with technical setup. Many thanks to Egon L. van den Broek for several valuable theoretical suggestions, to Blasimir Villa Rodriguez for practical discussions about concepts and careful proof-reading and to anonymous reviewers for helpful improvement suggestions.

References

1. Norman, D.A.: *The Psychology of Everyday Things*. Basic Book Inc., New York (1988)
2. Gibson, J.J.: *The theory of affordances*. In: Shaw, R., Bransford, J. (eds.) *Perceiving, Acting, and Knowing*, Erlbaum, Hillsdale NY (1977)
3. Gibson, J.J.: *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston (1979)
4. McGrenere, J., Ho, W.: *Affordances: Clarifying and evolving a concept*. In: *Proc. of Graphic Interface*, Montreal, Canada, pp. 179–186 (2000)
5. Oliver, M.: *The problem with affordance*. *E-Learning* 2(4), 402–413 (2005)
6. Norman, D.A.: *Affordances, conventions and design*. *Interactions*, 38–43 (1999)
7. Norman, D.A.: *Memory and Attention: an introduction to human information processing*. Wiley & Sons, London (1969)
8. Gaver, W.W.: *Technology affordances*. In: *Proc. of CHI*, New Orleans, U.S. (1991)
9. Hartson, H.R.: *Cognitive, physical, sensory and functional affordances in interaction design*. *Behaviour & Information Technology* 22(5), 315–338 (2003)
10. Vainio, L., Symes, E., Ellis, R., Tucker, M., Ottoboni, G.: *On the relations between action planning, object identification, and motor representations of observed actions and objects*. *Cognition* 108(2), 444–465 (2008)
11. Sheridan, J.G., Kortuem, G.: *Affordance-based design of physical interfaces for ubiquitous environments*. In: Youn, H.Y., Kim, M., Morikawa, H. (eds.) *UCS 2006*. LNCS, vol. 4239, pp. 183–199. Springer, Heidelberg (2006)
12. Luo, P., Eikman, E.A., Kealy, W., Qian, W.: *Analysis of a mammography teaching program based on an affordance design model*. *Academic Radiology* 13(12), 1542–1552 (2006)
13. Schegloff, E.A., Sacks, H.: *Opening up closings*. *Semiotica* 8(4), 289–327 (1973)
14. Bumann, H.: *Lexikon der Sprachwissenschaft*, 3rd edn. Kroener Verlag, Stuttgart (2002)
15. Cassell, J., Bickmore, T., Vilhjalmsen, H., Yan, H.: *More than just a pretty face: Affordances of embodiment*. In: *Proceedings of 2000 International Conference on Intelligent User Interfaces*, New Orleans (2000)
16. Holzinger, A.: *Usability engineering for software developers*. *Communications of the ACM* 48(1), 71–74 (2005)
17. Theune, M., van Schooten, B., op den Akker, R., Bosma, W., Hofs, D., Nijholt, A., Krahmer, E., van Hooijdonk, C., Marsi, E.: *Questions, pictures, answers: introducing pictures in question-answering systems*. In: *Proc. of the 10th. International Symposium on Social Communication*, Santiago de Cuba, Cuba (2007)

18. Mayer, R.E.: *Multimedia learning*. Cambridge, MA (2001)
19. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic media in computer science education; content complexity and learning performance: Is less more? *Educational Technology & Society* 11(1), 279–290 (2008)
20. op den Akker, R., Bunt, H., Keizer, S., van Schooten, B.: From question answering to spoken dialogue: Towards an information search assistant for interactive multi-modal information extraction. In: *Proc. of Interspeech*, Lissbon, Portugal (2005)
21. Wharton, C., Rieman, J., Lewis, C., Polson, P.: The cognitive walkthrough method: A practitioner's guide. In: Nielsen, J., Mack, R. (eds.) *Usability Inspection Methods*, pp. 105–140. Wiley, New York (1994)