Affordances and Cognitive Walkthrough for Analyzing Human-Virtual Human Interaction

Zsófia Ruttkay[1,2] and Rieks op den Akker[1]

[1] HMI, University of Twente, The Netherlands
[2] ITK, Pázmány Péter Catholic University, Hungary
zsofi@cs.utwente.nl

Abstract. This study investigates how the psychological notion of affordance, known from human computer interface design, can be adopted for the analysis and design of communication of a user with a Virtual Human (VH), as a novel interface. We take as starting point the original notion of affordance, used to describe the function of objects for humans. Then, we dwell on the human-computer interaction case when the object used by the human is (a piece of software in) the computer. In the next step, we look at human-human communication and identify actual and perceived affordances of the human body and mind. Then using the generic framework of affordances, we explain certain essential phenomena of human-human multimodal communication. Finally, we show how they carry over to the case of communicating with a 'designed human', that is an VH, whose human-like communication means may be augmented with ones reminiscent of the computer and fictive worlds. In the closing section we discuss and reformulate the method of cognitive walkthrough to make it applicable for evaluating the design of verbal and non-verbal interactive behaviour of VHs.

Keywords: Affordance theory, virtual humans, embodied conversational agents, cognitive walkthrough, HCI.

1 Introduction

James J. Gibson coined the term *affordance* [8] in order to describe the relationship between animals and their environment. Donald Norman introduced the new term *perceived affordance* to represent the relationship between users and artificial designed products [11].

Recently this notion, which has been in use as crucial concept to evaluate the design of technological products (including computer programs), has been used in a much broader sense in the community of researchers developing computer systems where the user communicates with virtual humans. Virtual humans (VHs) are synthetic characters produced by computational means, which are intended to look like and communicate as real people do [1]. The design and evaluation of VHs, as new generation in user interfaces, is a hot topic [13]. Cassell talked already years ago about the "affordances of the human body" [1], Marsella raised the issue whether nonverbal signals can be understood as (perceived) affordances [9], which idea was

A. Esposito et al. (Eds.): HH and HM Interaction 2007, LNAI 5042, pp. 90–106, 2008.

explored further by Ruttkay and Ten Hagen [16]. These initial and incomplete examples ask for a thorough analysis of this notion and its benefits for human-virtual human interaction (HVHI). The question is if we gain new insight by adopting the notion of affordance in the broader sense, as a useful novel reference framework to analyse HVHI. Or, are notions we are already familiar with, like non-verbal signals with meaning, equally suitable for the same analysis of phenomena and properties of this new type of user interface?

We claim for the first, affirmative answer: the concept of affordance, if taken in a broad sense, not only lends itself as a good frame of reference for human-VH communication, but makes its design and analysis transparent. It enables the researcher to treat human- and computer-related communications in the same way, and to use the traditional method of cognitive walkthrough of HCI for the novel case of HVHI.

Affordance theory has been used primarily in connection with the new, mediated communication means of computer technology. Whittaker [18] used the concept to confront the communication by mediated technologies, that is, the traditional means of HCI, and face to face communication. He did not extend the concept for phenomena in human-human interaction. Gaver [**7**] investigated the affordances of the physical environment, such as new technologies, for social interaction and emerging new behaviours. In robotics, learning affordances was addressed [11]. Affordance theory has been used as the basis for rapid generation and reusability of synthetic agents and objects [**3**]. The authors borrowed affordance theory to solve the AI problem of multiple representation of views of the same world in different agents' mind. They found the concept very useful, both for design and engineering multi-agent systems: "We conclude that Affordance Theory is an elegant solution to the problem of providing both rapid scenario development and the simulation of individual differences in perception, culture, and emotionality within the same agent architecture." They restricted the usage of the concept for describing the utilities of the (virtual and real) world for agents with different background. What they exploited from the theory is that the same object may offer different usage for different agents.

In [10] the authors compare Gibson's original definition with the use of the term by Norman. They also present a survey of the use of the concept in the HCI literature, more specifically in the annual CHI conference proceedings. Their conclusion is just the opposite of ours: "As the concept of affordances is used currently, it has marginal value because it lacks specific meaning."

In our paper we set out to grasp the meaning of affordance in human-computer interaction (HCI) and in particular, in HVHI. In the rest of the paper we will argue step by step. We take as starting point the original notion of affordance, used to describe the usage of objects by humans. Then, we dwell on the human-computer interaction case when the object used by the human is a (piece of software in) the computer. In the next chapter, we look at human-human communication and identify actual and perceived affordances of the human body and mind. Then using the generic framework of affordances, we explain certain essential phenomena of human-human multimodal communication. Finally, we show how they carry over to the case of communicating with a 'designed human', that is a VH, whose human-like communication means may be augmented with ones reminiscent of the computer and fictive worlds. We give an illustrative example from the case of conversation with an VH assistant for money retrieval. In the closing section we discuss and reformulate the method of cognitive walkthrough so as to make it applicable for evaluating the design of verbal and non-verbal interactive behaviour of VHs.

2 The Notion of Affordance in HCI

2.1 The Original Concept of Affordance

Originally, affordance was invented to describe how animals make use of objects in their environments [8, p. 127.]:

> "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. If a terrestrial surface is nearly horizontal (instead of slanted), nearly flat (instead of convex or concave), and sufficiently extended (relative to the size of the animal) and if its substance is rigid (relative to the weight of the animal), then the surface affords support... Note that the four properties listed - horizontal, flat, extended, and rigid - would be physical properties of a surface if they were measured with the scales and standard units used in physics. As an affordance of support for a species of animal, however, they have to be measured relative to the animal. They are unique for that animal. They are not just abstract physical properties. "

Note how in its original, most limited meaning affordance is relative to the 'user' of an object. This characteristic will be crucial in the subsequent generalizations too.

2.2 Perceived and Actual Affordance

What is the meaning of affordance in HCI? The notion of affordance was introduced in HCI by D. Norman [12]:

> "... the term affordance refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could be used. A chair affords ('is for') support and, therefore, affords sitting."

In their widely used course book "Human Computer Interaction", Alan Dix and others defines the concept in the following way [6, p. 135.]:

> "The psychological notion of affordance says that things may suggest by their shape and other attributes what you can do to them: a handle affords pulling or lifting, a button affords pushing. These affordances can be used when designing novel interaction elements. (…) Affordances are not intrinsic, but depend on the background and culture of users. Most computer scientists will click on an icon. This is not because they go around pushing pictures in art galleries, but because they have learned that this is an affordance of such objects in a computer domain.…

> Some psychologists argue that there are intrinsic properties, or affordances, of any visual object that suggest to us how they can be manipulated. The appearance of the object stimulates a familiarity with its behaviour. For example, the shape of a door handle can suggest how it should be manipulated to open a door, and a key on a keyboard suggests to us that it can be pushed. In the design of a graphical user interface, it is implied that a soft button used in a form's interface suggests it should be pushed (though it does not suggest how it is to be pushed via the mouse). Effective use of the affordances which exist for interface objects can enhance the familiarity of the interactive system. "

The above definitions adapt the original affordance concept by putting the human in the actor role, and the (designed) artefacts, especially computer systems in place of the objects of the natural environments. Moreover, the single notion is split into two related, but different notions of affordance (see also [16]):

- *perceived affordance*: what the object or system suggest the user can do or use it for;
- *actual affordance*: what the object or system can actually do/be used for.

The perceived affordance depends on two factors. First of all, on the appearance of the object or system. This is usually a visual appearance, but acoustic signals may also be used. E.g. the affordance "to read a message" may be indicated by an icon, and/or by an earcon, on the mobile phone or a computer screen. The appearance may be that of the object itself (e.g. the example of chair above), or may be a separate, designed signal (e.g. the exit signal above a door). Second, the mapping of the visual signal to the assumed function is done by the receiver, and is a result of learning. Things are in order if the perceived and actual affordances coincide. This requires, basically, that the user associates the perceived signal with the actual affordance. Such an association is often the result of learning. There are established, though may be in different ethnic or professional groups different, signals for certain actual affordances.

In HCI, the user interface designer is the one to care for a match between the perceived and the actual affordance. In general the appearance of the user interface should make clear to the user the 'tools' available in each situation, and these tools should be suited to perform each of the acts the user could possibly intend to do in a given situation. This could be considered as an important design principle for human-computer interfaces; a principle that we will call the *Affordance Requirement:*

> For every item, visible on the user interface at any time during the interaction with the user, it's perceived performance should match it's actual performance.

There is a strong relation between the above formulated notion of affordance with the requirements mentioned and Cognitive Walkthrough, a well-known method for evaluation of user interfaces. The very sense of a the Cognitive Walkthrough method

for evaluating a computer system is to check whether the user interface satisfies the design principles, one of these is the Affordance Requirement. We come back to this later.

2.3 Devices and Functions in HCI

What kind of affordances are common in HCI? For a careful analysis of devices and their function in HCI, we look at them from different aspects.

Realization: Physical versus abstract. In case of a computer, we must differentiate between physical input and periphery devices as physical affordances, and abstract 'devices' which form part of the software, and are to perform specific tasks. Examples of physical devices: the keyboard, the mouse, a printer attached to the pc, … Examples of abstract devices (all depicted on the display): a trash-bin, a screen-locker, a mailbox.

In the first case, using/manipulating the physical device results in some action by the device, like entering a character, locating a point on the screen. (We'll return to the printer later.) In the second case, there are conventional ways to activate the abstract devices, either by dragging and dropping some 'object' the device has to operate on (e.g. dragging a file to the dustbin, to be thrown away), or to click the device to activate the function (e.g. locking the screen by clicking the lock-screen icon). In case of abstract devices thus, one uses simple physical devices (most often, the mouse, or a specific key combination of the keyboard) to activate them.

Range of functions: Simple (or basic) versus complex. The printer and the mailbox are complex devices in the sense that they make possible a set of functions, each to be activated by some sub-part of the entire complex device. E.g. on the printer, the printing options can be specified, a page can be printed, or a new paper can be fed. A mailbox usually stands for a mail program, allowing to send/receive/store and sort mails.

Range of objects/targets they (may) act on: Given versus to be chosen. The target of activation may be specified too (the file to be selected which needs to be deleted, printed, …). This is not unlike with the affordances in the real world. While a lamp switch will always act on a single lamp, a knife as an affordance to cut must be applied to some object. The range of possible targets is, however, constrained: one should not try to cut a piece of metal or stone with a knife. This analogy also applies for HCI, where the parameters of files may be decisive if it is a proper object for an affordance, e.g. printing can be applied only to specific types of 'printable' documents.

There may be other side-conditions to activate some abstract devices, like time, availability of certain resources. So some abstract devices are applicable in certain contexts only. The context can be that a proper target has been selected (in an implicit way). Hence an abstract device may be context dependent or independent. A complex physical device may have some context dependent behaviour: the print function may be active only if there is printing paper available.

Another observation is that an abstract device may be linked directly or indirectly to a physical device, think of the example of a printer drive as abstract device and a printer as a physical device.

3 Affordances in Human-Human Interaction

3.1 Affordances of the Speaker and the Listener

The original context of affordances was the utility of objects for humans (and animals). In HCI, the 'objects of the world' was replaced by physical and abstract devices of computers. In both cases, the intelligent human is the one to act and use the affordances of the (unintelligent, non-reactive) world. In case of computer affordances, the 'dead objects' though may show some reactive and dynamic behaviour, limited to indicating applicability in a given context.

Now let's have a look how affordances are attributed to the human body by J. Cassell [1]: "Only humans can communicate using language and carry on conversations with one another. And skills of conversation have developed in humans in such a way as to exploit all of the unique *affordances of the human body*." Affordances seem to be essential for human language usage, and in more general, to interactive behaviour. Examples are multimodal gestures, like an open hand, or looking at the partner, which signal that one has given turn, and thus is can be talked to now.

At a first glance it seems that in the conversational context the term 'affordance' is used in a different sense. What is to be achieved is certain *behaviour* of the conversant, who isn't an object that a user is supposed to do something with. But the behaviour of the other is a necessary condition for the user to perform some action: one speaks only if there is a listener around. And the listener does signal that he is not a deaf person being present, but one listening to and understanding what the other is saying. Hence there is signalling of a *function* the speaker may exploit, albeit this function is different from that of using an object as a tool: it is to manage communication.

When talking about affordances in human-human interaction, there are major differences, with respect to human-world or human-computer interaction:

- It takes place between two (or more), in facilities and capabilities (by and large) *equal parties*.
- For a successful conversation, constant *co-operation by the speaker and listener* is needed.
- A human person has an *arsenal of physical and mental affordances* (see below) to be used for communication, and they adapt to the situation and partner in using the most appropriate one.
- The correspondence between perceived and actual physical and mental affordances is more intricate, as the *mapping is many to many*, and it highly *depends on culture and other factors*.

3.2 Comparison of Affordances in Human-Object, Human-Computer and Human-Human Interaction

The perception and usage of affordances was a relatively simple task in the previous two cases: the human had to recognize an affordance and make use of its unique function. (He had to discover the switch next to the door, and use it if he wanted to have the light on.) In human-human communication, one of the parties, the speaker does take the initiative with addressing the other, the listener, assuming that the partner is a

receiver of what he is to say. However, it highly depends on the partner if he will actually function as a receiver of the words uttered by the speaker. Problems may arise at different levels: it is not sure if the partner hears him, understands him, is interested in what he is saying. In affordance terminology, the listener may not have (permanently or temporarily) all the physical and mental affordances needed altogether to function as a listener. In human-human communication, it is thus common to give feedback on the availability of simple or complex, mental and physical affordances.

In concrete, talking to somebody may invoke very different reactions from him/her, such as:

- No sign from the listener, as if deaf, or not noticing that he is talked to;
- Some sign (e.g. puzzled facial expression) from the listener indicating that he has not understood what the speaker was telling.
- Answer or action according to what the speaker has told.

When a speaker (S) is talking, it is assumed that the listener (L) can hear and understand S, and that L is interested and listening to S. Thus for a successful conversation one has to take into account the affordances used by S (English speech) and the affordances available at the moment for L (hearing, understanding English, etc.). The multitude of affordances, their temporal availability and the symmetrical role of the communicating partners are the major differences and cause of complexity of the affordance concept in human-human interaction. In Table 1, we give a comparison of the concept in the three domains.

Below we look at the different aspects in detail.

Physical vs mental affordance. *A physical affordance* is some low-level functionality of a part or organ of the body, e.g. the ability to move limbs, to make facial expressions, to articulate voices, to sense sound or to see. A *mental affordance* is a capability of interpreting signals produced by a physical affordance, e.g. talking and understanding a language, interpreting facial expressions. Note that mental affordances assume the existence of certain physical affordances: to be able to interpret facial expressions, the vision system must be functioning. (We will not deviate here to a discussion of the extent to which a mental affordance is the sum of low-level physical affordances, if one allows that all low-level bodily functions are physical affordances.) Moreover, it is a widely shared opinion that most of the mental affordances are learnt, and are culture-specific. In the VH community J. Cassell talks about the 'affordances of the human body' in the sense of physical affordances, as 'faculties' or 'devices' for (multimodal) communication [1].

Role of affordances in human-human communication. In the physical world, or in traditional HCI, the affordance was always associated with an object, from the point of view of the human user. Hence there the role is asymmetric. In human-human communication, the communicating partners play, in principle, a symmetrical role. A person's affordance of hearing and English language understanding are valuable for somebody else who has the matching affordances of the ability to talk and speak English. One may speak several languages, thus having a multitude of affordances, each valuable only for certain people, namely those speaking a given language. Hence

Table 1. Comparison of aspects of the concept affordance in human-object, human-computer and human-human interaction. In the last columns, the affordance is discussed from the speaker's (S) point of view. Note however, that S has own physical and mental affordances essential to initiate communication.

Aspect\Type of interaction	**Human-object**	**Human-computer**	**Human-human**
The 'user' of the affordance	A human	A human	Human speaker (S) or listener (L)
The affordance	The non-responsive object	The static physical or abstract device, may be with a few and well-understood possible states (e.g. with a printer: paper jam, ink problem, or ready to print)	An active human (L), responsive, with a wide range of possible states related to availability of actual physical and mental resources, but also individual deviations
Actual affordance	Physical, one per object	Physical or abstract, few per object	Physical or mental, several per human
Perceived affordance (signal)	The object itself, or a static designed visual/acoustic sign to indicate function	The object itself, or a designed visual/acoustic sign, with a few parameters to indicate context	Multimodal signals produces by L in a non-deterministic way during the conversation as feedback.
Signalling lack of actual affordance	Rarely (e.g. red light indicating problem with device)	Sometimes (e.g. broken connection)	Often signalled
Decoding the actual affordance from perceived one	Based on learning, experience and agreed design protocols	Abstract affordances are not always easy to decode, due to lack of or misuse of design protocols. Mapping is one to one, though.	Coding (by L) decoding (by S) protocols are complex, may differ, and many to many mappings often exist between signal and function

affordances of a speaker and a listener complement each other. Note that this relative nature of the affordance concept was present in how Gibson used the term, then for animals exploiting objects of the environment: a flat surface may be a suitable seat for one animal, and not so for another one. But the difference is that in the animal-object relationship Gibson talked about different species of animals with very different

physical and physiological characteristics, while in the case of human-human communication the relativity of an affordance is caused by the diversity in learnt, mental affordances over the (reatively-speaking) physically uniform humans.

We can talk about symmetry in an another sense too, namely that two conversing humans posses, by and large, an identical set of physical and mental affordances, and they use different – but not disjoint – subsets depending on who is the speaker and who is the listener.

Perceiving affordances of a human. How to find out if a person has the affordance of understanding spoken English? A straightforward way is to ask - but if the person is deaf, or does not understand English, he or she will not be able to answer. On the other hand, in real life communication, the person spoken to does in general signal back if they have heard and grasped what he was told, also stating in an indirect way that he can hear and understand English [13]. In human-human communication there are two layers, both of which may make use of verbal and/or nonverbal modalities [17]:

- the transfer of some information content;
- a meta-level signalling, which provides feedback (among other things) about the availability and proper functioning of the devices needed to organize the conversation and decode the message.

Affordances for dialogue control. A conversation requires the coordination of the division of floor: who is the speaker and who is listening in each moment. The low-level bodily affordances (eye gaze, posture, speech characteristics) are used to signal such information as turn to be given, turn asked for, turn taken and kept, or listening to the speaker. Hence the multi-modal signals can be seen as visualizations of some 'state' the human is in a given moment: listening, talking to the partner, finishing his/her speech, recalling data while talking. In these states, one utilizes different sets of affordances. The different states can be thus associated with the affordances available for the conversants. E.g. if someone is in 'talking' state, the partner should not interrupt, which would be appropriate only if the speaker stops talking and indicates that now the 'listening' affordances are available. As well as signalling one's own affordances being active, one may also signal to the partner a particular affordance they are required to use. E.g. gazing at the partner signals that one has finished talking and can be talked to from now on. Further on, after some time if the partner has not started talking, pointing at him with open palms upwards as a sign that he should take the turn, that is, use the relevant physical and mental affordances. So in human-human communication we can identify affordances and differentiate the two types introduced by Norman, as follows: Actual affordances are the arsenal of physical and mental faculties of a human to communicate. In order to perceive an affordance, some speech and/or nonverbal signals may be produced that the human has some actual affordances operating, or just about to use, or not using, or having problems with using it. Note that in the extreme, one may not produce any signal to indicate the existence of an affordance, e.g. listening to somebody without any feedback, with a poker face.

As we see, the affordance concept becomes more complex because of the huge set of potential affordances, of which only some are available at each moment. Also a further feature is that verbal and nonverbal signaling is used to indicate not only (in the original view, static) availability of an affordance, but also temporary problems with (or lack of) an affordance.

The decoding of the signal by the partner as intended by the signaling person. While in case of objects and computers as affordances, there may be some culture-dependent protocol for the design of the visualization of the affordance (e.g. design style of the 60-ies, or Swedish design for lamp switches), in case of human-human communication the mapping between affordances and signals is far more complicated. The cultural and social background of the listener plays a role in how he decodes signals, that is, what affordance he perceives. Moreover, there is a variety from time to time how a signals is (not) used, according to the static or dynamic characteristics of the speaker. E.g. an introverted person may use fewer and less articulate facial expressions for back-channelling a conversation; a certain eyebrow movement may be the idiosyncratic signal in case of a given person of boredom, sadness may alter the signaling of back-channeling, etc.

Separation of signal, affordance and target of action. If we look at affordances of the real world, there we find two kinds: ones where the target of operation is unique, the object having the affordance itself (e.g. to open a door, the handle needs to be pushed – the affordance has a single function, to open the given door), and ones which may be applied to achieve certain goals (e.g. an iron bar may be used to break in all kinds of doors, windows, using the affordance of the iron bar 'breaking firm objects'.) In the designed world we already stated that the signal may be separate form the object itself (e.g. sign of exit). In human-human interaction (HHI), the discourse regulation affordances are signaled by some organs and body parts, which are not the target of the action (the hand signals but is not the object the partner is expected to operate on). If we think of the analogy of opening a door, then the task can be the collaborative interaction of two communicating partners. But on the interaction level, they use the communicational affordances which make the information exchange possible (by speech, or other means) which may lead to solving the problem. Hence affordances in HHI are on a meta-level, to assure successful communication, and not on the level of solving some concrete task. In this respect the function of affordances of HHI is similar to the function of some of the affordances in HCI, namely those which indicate for the user to 'take a step' from the allowed ones, e.g. by typing in a command, or confirming an action to be performed by the system.

4 Affordances for Human-Virtual Human Interaction

4.1 Virtual Humans as Design Products

With human-virtual human interaction, we arrive back to a special type of human-computer interaction, that is, to a domain where everything offered for the user is the result of conscious design decisions (or unconscious mistakes). In order to explore the merit of the affordance concept for H-VH communication, we first turn to Daniel C. Dennett's theory about how people try to explain why things and phenomena are the way they experience them. He distinguished three stances that we can take towards these phenomena.

Physical stance. Clarification is based on laws of nature, e. g. this stone falls because of its mass and the law of attraction. Causes are physical causes.

Design stance. Clarification is based on the function it has obtained by the technical designer. All arts and technology, handcraft as well as machines work because of their design by humans. They use laws of physics and express new combinations of natural laws and qualities in order to make something happen that suits humans goals.

Intentional stance. Clarification based on motives or intention or goals that the system or thing itself has. Some people say that computers or robots have intentions (as humans have).

Note that often it depends on the person how he/she looks at a phenomenon: from a physical,. design or intentional stance. For instance, a mechanical clock works by exploiting physical laws, and the motion of its parts can be explained as a physical process. However, the clock, as a device is a designed artifact. As such it can be (and most naturally it is) looked at from a design stance. A person who has never seen a clock, may take an intentional stance and say that "the big hand chases the small hand", or that a bird flies out every now and then to look around.

Then, there is the important distinction between spontaneous, immediate reaction versus using something as a tool. The fact that my behaviour has some particular effect on someone else, because he interprets this behaviour in a certain way, doesn't imply that I use this behaviour deliberately to bring about this responsive behaviour. Natural (and non-verbal) language use is different in this respect from the use of a tool as a means to reach some goal.

People usually indicate in one way or another their feelings or mood, unless they make efforts to hide them, due to social 'display rules' or personal reasons. Others thus usually can notice - from the way how someone looks, stands or walks – one's emotion, mood and physical state. These are spontaneous expressions, not signals deliberately produced in order to express these feelings.

In a computer application with a VH, human behaviour is used in a consciously designed way, in order to reach some effects on the user. What was natural behaviour in the first place now, from the system designer's point of view, becomes a 'tool' to bring about desired effects in interaction with the human user of a system. E.g. in order to make the user answer, the VH should turn towards the user at the end of his speech, and keep gazing at the user. Looking at this from the user's point of view, he takes the intentional stance, and attributes intention of 'waiting for an answer' to the VH. The VHs are 'good' if they fool the user to believe that the VH acts and behaves intentionally, and really understands what the user is saying.

The above mentioned usage of 'natural' signaling by VHs, however, can be seen just as signals in HCI to inform the user what may or may not be allowed in a given situation.

It is the perspective of the design stance by Dennett which links the (spontaneous) natural human behaviour and the consciously chosen communicational protocol of HCI, the signals used among humans and the specific ones used in HCI. It's not uncommon nowadays to talk about 'facial display' the way people show for an outside observer how they want to be seen. From this stance we see verbal and non-verbal behaviour as presenting signs in order to signal what the subject self is doing or what the subject self expects the observer to do. The use of the notion of affordance for discourse functions in natural behaviour thus gives us a more clear picture of the meaning it already had in classical HCI literature.

Further on, in H-VH communication, the 'natural' affordances and their signaling protocols may be used interwoven with the unnatural, designed protocols of HCI. In other words, there may be unrealistic capabilities and signalling mechanisms of VH, which can be more efficient than the natural ones. Such an example is when, in a scene where multiple VHs are present, the one to be addressed (the listener in a given moment) is indicated by a red arrow above the head, instead of or in addition to the normal natural signaling which may be difficult to produce or notice from a distance, or due to limitations in animation and rendering. But one may think of a richer repertoire of bodily affordances. E.g. if speech not understood properly, the VH's ears could grow big, indicating the problem with catching words. This is similar to what people do when they enlarge the ears by putting the two open palms behind. While such a non-realistic capability is accepted in animation films, mixing of real and fictional, or human and computer interaction protocols is yet an unexplored field for VHs. The question, basically, reduces to "what extent is a virtual human looked upon as a real one?" The answer surely depends on the application domain.

4.2 An Example

Below we illustrate the discussed aspects of affordances in HVHI. The example assumes an internet banking site which is supervised by a VH as assistant. In our example we assume that the interaction takes place by the traditional question-answer protocol. Only if problems arise, or the user has not taken action for a long time or asked for help explicitly, the VH intervenes, as in the example. The human user (U) has reached the point where his account number is to be given. For the first time in the interaction, there is no input from him for 10 seconds. Then the VH appears:

VH: "Give your bank account number!"
U: No reaction for another 10 seconds.

Let's have a careful look at the possible causes of the silence:

1. Listener is deaf.
2. L does not speak English.
3. L did not catch the words.
4. L did not realize he was addressed.
5. L expects more polite treatment.
6. L does not know how to give the required number.
7. L does not have a bank account
8. L cannot answer now, as he is busy with talking to somebody.
9. L has his right hand in cast, is busy with figuring out the numbers with a left finger,
10. The user told his bank account, as an answer for the speech of the VH.
11. L had left earlier.

The nature of the cause of the user not answering is different in the above cases, related to problem with

- the physical affordance as hearing, needed to be addressed (1, 3),
- the mental affordance of understanding English (2)
- mental processing capacity (8),

- bodily affordance (9)
- perceiving an affordance (4, 6)
- the perceived and actual affordance of the VH (10),

So in such a simple case there can be at least 6 kinds of problems with affordances. (Of course, the cause of delay or no reaction can be other than problems with affordances, like lacking the information required for the reaction (7), not trusting the VH because of the ordering tone (5), or not being present at all (11).) Note that the problem with 3 may be attributed also to VH (e.g. speaking in a robot-like synthetic voice, difficult to understand), not only to H (not hearing well, or not mastering English). But in this case it may be the context (e.g. noisy street scene) which made speech as an affordance inappropriate. In case of 4. and 6. there is a problem with associating a signaled affordance (talking to somebody, indicating on a screen that digits are to be typed in) and the actual affordance. It may be the user who did not do the right mapping of the signaled and actual affordance, but it may also be that the signaled affordance was poor: the VH who did not signal clearly enough that he was addressing the user or it is not indicated at all on the screen that some digits are to be typed in by the user. If the latter, it is clearly a design mistake.

When designing the dialogue with the VH, it should be investigated how the above cases need to be dealt with. First of all, the design of the VH and the system should be such that many of the possible causes get eliminated. E.g. the VH should address the U in a noticeable way, by turning towards him, changing posture. The voice quality, loudness, speech tempo should be chosen such that it is easy to catch the words, the noise level of the environment should be considered.

Some other causes (8, 9, 11) can be excluded if the VH has visual perception capabilities, but this is usually not the case. A common, related problem with VH is that more is expected from the VH than it is capable of, in terms of natural communication. In our case, it is assumed that the speaking VH can also hear and understand language, which may not be true. Such false expectations can be countered by telling explicitly at the beginning, or in the commands, what input modalities are to be used. That is, in our case "Type in your bank account number."

So the dialogue should be designed in such a way that no possible cause of miscommunication remains ignored. Which is a challenging task, as in a given situation it is only one of the 10 possible problems which causes no answer by the user. There can be different AI techniques exploited to make the best guess, like single user modeling, tuning probabilistic transitions based on statistical analysis of performance. Another, complementary approach could be to use more modalities for signaling affordances (e.g. blinking space where the numbers need to be typed in, or the VH pointing at the screen location while asking for the input).

5 Cognitive Walk through to Analyse HVHI

Cognitive walkthrough was introduced to underpin the informal and subjective walkthrough technique with psychological theory. We present the revised version of this method for evaluating the design of the human computer interface as it has been developed more recently to make it accessible for system designers [6]. We will discuss

here the relevance of this method for evaluating interactive behaviour, especially with respect to possible communication failures. Can these failures be avoided by a cognitive walkthrough of the interactive behaviour of the VH? Cognitive walkthrough is presented as an evaluation method to check whether the design of the interactive behaviour satisfies the Affordance Requirement. In the above example, this was happening in an unsystematic way already.

We first give the formulation of the method in terms of 'classical' interface design. Before you can start a cognitive walkthrough evaluation you need:

1. A detailed description of the system (prototype)
2. A description of the task the user is to perform with the help of the system
3. A complete, written list of the actions needed to complete the task with the system.
4. An indication of who the users are and what kind of experience and knowledge the evaluators can assume about them.

As you see it is rather indicative and not very precisely specified. And it is clear that the method doesn't apply without any modifications to the analysis of affordance issues in interactive behaviour using VHs. Cognitive walkthrough is the method in which the evaluator goes through the action sequence mentioned in item 3 above to "critique the system and tell a believable story about its usability" [6]. This is done by systematically trying to answer the sets of questions below in each situation. We, again, first, follow the formulation in terms of classical human computer interface elements.

1. Will users be trying to produce whatever effect an action has? Are the assumptions about what task the action is supporting correct, given the users' experience and knowledge up to this point in the interaction?
2. Will users be able to notice that the correct action is available? Will users see the button or menu item, for example, by which the next action is actually achieved by the system?
3. Once users find the correct action at the interface, will they know that it is the right one for the effect they are trying to produce? This complements the previous question. It is one thing for a button or item to be visible, but will the users know that if it is the one they are looking for to complete their task?
4. After the action is taken, will users understand the feedback they get? Assuming the users did the correct action, will they notice that?

The group 2 and 4 of the above questions can be reformulated such that they are appropriate for a system where the interface is a VH, and where the traditional interactio means of HCI, like buttons and menus are replaced by natural human interaction facilities, i.e. using natural language supported by non-verbal conversational gestures. Actually, this was the approach taken to trace places and causes of problems occurring with two implemented prototype systems at the VH Workshop [9]. Below are the questions adapted to test the design of the communication with a VH:

1R Will the user be aware of what they can do to achieve a certain task: to talk to several (all) VHs on the screen, to use gestures, gaze and head movement as those are perceived by the VHs?

2R Will users notice (when) they need to give an answer? Will they notice whom they may address (who is listening)?

4R Will there be a feedback to acknowledge the (natural, may be multi-modal) answer given by the user? Does the feedback indicate for the user if his/her action was correct? (In case of natural communication, this distinction means if something 'syntactically correct' was said and thus properly parsed, or something was said which is (one of the) expected answers in such a situation.)

Note that nothing prevents us from analysing systems with the method of cognitive walkthrough where natural communication takes place with a VH, interwoven with communication using elements of traditional HCI techniques. One has to keep in mind that 'signal' may be a natural communicational signal as well as a signal used in traditional HCI methods. Even more, the two types of signaling may be mixed. E.g. in a scene where several VHs are present, the one to be addressed may be indicated by an arrow, or blinking appearance.

Finally we mention that cognitive walkthrough is not the method to evaluate how efficient the user or the system is to solve some task. Think of two systems, one which allows to delete files one by one only, the other which allows to delete a selected set at once. Both can be well or badly designed, from a HCI point of view. Which has nothing to do with the fact that the 2nd system per se offers more functionality. What interface problem is, to be spotted by cognitive walkthrough, if a user in the 2nd case remains deleting a big number of files one by one, taking no notice of a functionality. Such problems as not using all the functionality of a system, or not using it efficiently could be also spotted by automatic methods, similar to ones used in the analysis of complex systems.

6 Conclusion

We have extended the use of the term affordance for the context of interaction with real and virtual humans, based on some abstract correspondence with the situation where the use of the word was already established and obtained a meaning: in traditional technology of designed objects (tools, devices and computer). From the design stance we can see that there is an aspect in human communicative behaviour - verbal as well as non-verbal - that structurally resembles certain communication means and their functions of traditional HCI. Based on the ideas of intentional and design stance we tried to unravel this deeper understanding. We identified the Affordance Requirement and reformulated the evaluation strategy known as cognitive walkthrough so as to make it applicable as an evaluation strategy for the design of the interactive behaviour of VHs. The proposed evaluation method seems particularly suitable for discovering potential flaws in this behaviour that may lead to communication failures. This aspects motivated S. Marsella to raise the issue, and later on, make empirical evaluations of different ECA systems on a 'cognitive walkthrough' basis [9].

We believe that a common terminology of systems with traditional and human-like interfaces gives a new insight to the latter systems, which have been compared to real human performance and evaluated from the point of their (subjective and objective) effect on the user. In our opinion, the generalization of affordances for human-human

communication helps to look at h-h (and thus, h-virtual human) communication from an objective, design and performance perspective; makes the effect of the characteristics of the user, the computer and operational environment explicit, and helps to bridge the gap between human-computer and human-virtual human interaction. This new point of view allows the adaptation of evaluation techniques well-known in traditional HCI and system evaluation, such as cognitive walkthrough or automatic runtime evaluation.

Future work, based on the clarified concept of affordances in human-virtual human interaction, could be a formal framework where design principles are formulated and can be checked systematically. Also alternative perceived affordances for a single actual affordance, both in the design stage and as choices available could be handled, depending on the physical environment (level of noise) , the implementation constraints (speech input available or not) and last but not least, the user group. In [16] we made a first step in this direction.

We finally come back to a basic principle of interface design, that we encountered earlier. "What you must not do is depict a real-world object in a context where its normal affordances do not work!" ([6] p. 217) Does this imply that we *must not* make computer interfaces that give the user the impression that our computer systems with a virtual human as interface understand 'natural' language when they actually do not? This question would lead us to the polemy of how appropriate it is to use virtual humans, which may get attributed with all the mental affordances of humans.

Acknowledgements

We are grateful to Herman Koppelman for pointing at the literature on design and evaluation, and for Paul ten Hagen and Dirk Heylen for discussing ideas and commenting on earlier versions of the paper. We are indebted for comments from the anonymous reviewers. This work was partially carried out in the framework of the COST2102 action . This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

References

1. Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): Embodied Conversational Agents. MIT Press, Cambridge (2000)
2. Cassell, J., Bickmore, T.W., Vilhjálmsson, H.H., Yan, H.: More than just a pretty face: affordances of embodiment. In: Proc. of Intelligent User Interfaces, pp. 52–59 (2000)
3. Cornwell, J., O'Brien, K., Silverman, B.G., Toth, J.: Affordance theory for improving the rapid generation, composability, and reusability of synthetic agents and objects. In: Proc. of 12th Conf. on Behavior Representation in Modeling and Simulation (2003)
4. Dennett, D.C.: True believers: The intentional strategy and why it works. In: Heath, A.F. (ed.) Scientific Explanations: Papers based on Herbert Spencer Lectures Given in the University of Oxford, Reprinted in The Nature of Consciousness (David Rosenthal, ed.) (1981)
5. Dennett, D.C.: The Intentional Stance. MIT Press, Cambridge (1987)

6. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human Computer Interaction. Prentice Hall, Englewood Cliffs (2004)
7. Gaver, W.: Affordances for interaction: The social is material for design. Ecological Psychology 8(2), 112–129 (1996)
8. Gibson, J.J.: The ecological approach to visual perception. Houghton, Mifflin, Boston (1979), http://www.alamut.com/notebooks/a/affordances.html
9. Personal communication at the AAMAS 2004 ws 'Balanced Perception and Action in ECAs' (2004)
10. McGrenere, J., Ho, W.: Affordance: Clarifying and Evolving a Concept. In: Proceedings of Graphics Interface, Montreal, Canada, pp. 179–186 (2000)
11. Murphy, R.R.: Case studies of applying Gibson's ecological approach to mobile robots. IEEE Transactions on Systems, Man, and Cybernetics, Part A 29(1), 105–111 (1999)
12. Norman, D.: The psychology of everyday things, Basic Books, New York(1988), http://www.jnd.org/dn.mss/affordances_and_desi.html
13. Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M.: Engagement Capabilities of ECAs. In: AAMAS 2005 Ws on Creating Bonds with ECAs, Utrecht (2005)
14. Ruttkay, Zs., Pelachaud, C. (eds.): From Brows to Trust: Evaluating Embodied Conversational Agents. Kluwer, Dordrecht (2004)
15. Reeves, B., Nass, C.: The Media Equation. How People Treat Computers, Television and New Media Like Real People and Places. CSLI Publication Cambridge University Press, Cambridge (1998)
16. Ruttkay, Zs., Ten Hagen, P.: Reactive Monologues Modeling Refinement and Variations of Interaction Protocols of ECAs. In: Proc. of AAMAS 2004, Workshop on Balanced Perception and Action in ECAs, New York (2004)
17. Thórisson, K.R.: Natural Turn-Taking Needs No Manual. In: Granström, B., House, D., Karlsson, I. (eds.) Multimodality in Language and Speech Systems, pp. 173–207. Kluwer Academic Publishers, Dordrecht (2002)
18. Whittaker, S.: Theories and Methods in Mediated Communication. In: Graesser, A., Gernsbacher, M., Goldman, S. (eds.) The Handbook of Discourse Processes, pp. 243–286. Erlbaum, NJ (2002)