Intuition as Instinctive Dialogue

Daniel Sonntag

German Research Center for Artificial Intelligence
66123 Saarbrücken, Germany
sonntag@dfki.de

Abstract. A multimodal dialogue system which answers user questions in natural speech presents one of the main achievements of contemporary interaction-based AI technology. To allow for an intuitive, multimodal, task-based dialogue, the following must be employed: more than explicit models of the discourse of the interaction, the available information material, the domain of interest, the task, and/or models of a user or user group. The fact that humans adapt their dialogue behaviour over time according to their dialogue partners' knowledge, attitude, and competence poses the question for us what the influence of intuition in this natural human communication behaviour might be. A concrete environment, where an intuition model extends a sensory-based modelling of instincts can be used and should help us to assess the significance of intuition in multimodal dialogue. We will explain the relevant concepts and references for self-study and offer a specific starting point of thinking about *intuition* as a recommendation to implement complex interaction systems with intuitive capabilities. We hope this chapter proposes avenues for future research to formalise the concept of intuition in technical albeit human-centred AI systems.

1 Introduction

Artificial Intelligence (AI) helps to solve fundamental problems of human computer interaction (HCI) technology. When humans converse with each other, they utilise many input and output modalities in order to interact. These include gestures, or mimicry (including drawings or written language, for example), which belong to non-verbal communication. The verbal communication mode is the spoken language. Some modes of communication are more efficient or effective for certain tasks or contexts. For example, a mobile user interface could be addressed by spoken language in contexts where someone is already busy with his hands and eyes (for example, while driving) or simply to avoid tedious text input modes on mobile devices such as smartphones. AI techniques [1] can be used to model this complex interaction behaviour and Machine Learning (e.g., see [2]) plays a significant role in the modelling of content information over time.

We are almost certain that multimodal dialogue-based communication with machines will become one of the most influential AI applications of the future. Who wouldn't like to speak freely to computers and ask questions about clicked or pointed items? Especially when the questions can be answered in real-time with the help of search engines on the World Wide Web or other information repositories. Eventually, the role of multimodal dialogue systems may shift from merely performance enhancers (voice input is fast and convenient on mobile devices) toward guides, educational tutors, or adaptable interfaces in ambient intelligence environments where electronic agents are sensitive and responsive to the presence of users (also cf. [3]).

Multimodal dialogue systems as intelligent user interfaces (IUIs) may be understood as human-machine interfaces that aim towards improving the efficiency, effectiveness, and naturalness of human-machine interaction. [4] argued that human abilities should be amplified, not impended, by using computers. In order to implement these properties, explicit models of the discourse of the interaction, the available information material, the domain of interest, the task, and/or models of a user or user group have to be employed. But that's not everything. Dialogue-based interaction technology, and HCIs in general, still have plenty of room for improvements. For example, dialogue systems are very limited to user adaptation or the adaptation to special dialogue situations. Humans, however, adapt their dialogue behaviour over time according to their dialogue partners' knowledge, attitude, and competence. This is possible because humans' abilities also include (1) the emotions that are expressed and perceived in natural human-human communication, (2) the instinctive actions and reactions a human dialogue participant performs, and (3) the metacognitive and self-reflective (introspective) abilities of a human dialogue participant to cleverly reason about the actions she or he takes [5].

Intuition is widely understood as non-perceptual input to the decision process: *in philosophy, [intuition is] the power of obtaining knowledge that cannot be acquired either by inference or observation, by reason or experience.* (Encyclopdia Britannica). In this chapter, we ask two things: first, how does intuition influence natural human communication and multimodal dialogue, and second, how can intuition be modelled and used in a complex (dialogue-based) interaction system? We think that by providing a concrete environment in which an intuition model extends the sensory-based modelling of instincts to be used (multimodal sensors reveal current state information which triggers direct reaction rules), we can assess the significance of intuition in multimodal dialogue and come one step closer to capturing and integrating the concept of intuition into a working HCI system.

In section 2 we will introduce the application environment, i.e., the multimodal dialogue systems for implementing intuition in dialogue. A dialogue example describes dialogue topics, topic shifts, and an intuitive development of a common interest. Section 3 discusses dialogue adaptivity based on a context model which includes user modelling, and dialogue constraints and obligations. Section 4 provides a definition of instinctive dialogues based on multimodal sensory input and the recognition of moods and emotions. Section 5 provides the reader with our main contribution, a model for implementing intuitive dialogue which includes a self-reflective AI model. In section 6 we come to a conclusion.

2 Multimodal Dialogue Environment

Natural Language Processing (NLP) is a wide sphere.[1] NLP includes the processing of spoken or written language. *Computational Linguistics* is a related field; it is a discipline with its roots in linguistics and computer science and is concerned with the computational aspects of the human language faculty. Dialogue systems research is a sub-discipline of computational linguistics and works at allowing users to speak to computers in natural language. Spoken dialogue technology (as a result of dialogue systems research) additionally includes related engineering disciplines such as automatic speech recognition [7] and is the key to the conversational (intelligent) user interface, as pointed out in [8, 9].[2]

Multimodal dialogue systems allow dialogical inputs and outputs in more than just one modality and go beyond the capabilities of text-based Internet search engines or speech-based dialogue systems. Depending on the specific context, the best input and output modalities can be selected and combined. They belong to the most advanced intelligent user interfaces [11] in comparison to text-based search engines.[3] Some prominent end-to-end multimodal dialogue system are Janus [13], Verbmobil [14], Galaxy and Darpa Communicator [15–17], Smartkom [18] and SmartWeb [19–21]. Figure 1 shows different input and output modalities of interfaces used in multimodal dialogue systems as applications of multimodal dialogue in mobile or ambient intelligence environments.

We will start by explaining the interaction with multimodal dialogue systems according to Julia Freeman's *specious dialogue* (specious means apparently good or right though lacking real merit) artwork.[4] In her concept, dialogues are embodiments and consist of pairs of movable, sculptural forms (like physical dialogue agents, cf. the reproduced illustration in figure 1, left). They can play a multitude of roles such as lurking in corners or shouting at visitors, or "... they will expect to be touched or moved in some way, at the very least they will want to be listened or spoken to" (Freeman). Interestingly, this simple conception leads us to two very dominant roles of multimodal dialogue systems. First, the dominant interaction mode of spoken language, and second, the dominant social role of using haptics and touch to establish a relationship with natural and artificial things. The tradition of spoken dialogue technology (figure 1, upper left) reflects this dominant interaction role. Telephone-based dialogue systems, where the automatic speech recognition (ASR) phase plays the major role were the first (monomodal) systems which were invented for real application scenarios, e.g., a travel agent speaking by phone with a customer in a specific domain of

[1] [6] discusses a comprehensive set of some topics included in this sphere.

[2] Also see [10] for an introduction to relevant topics, such as dialogue modelling, data analysis, dialogue corpus annotation, and annotation tools.

[3] An introduction to multimodal dialogue processing can be found in [12].

[4] http://www.translatingnature.org

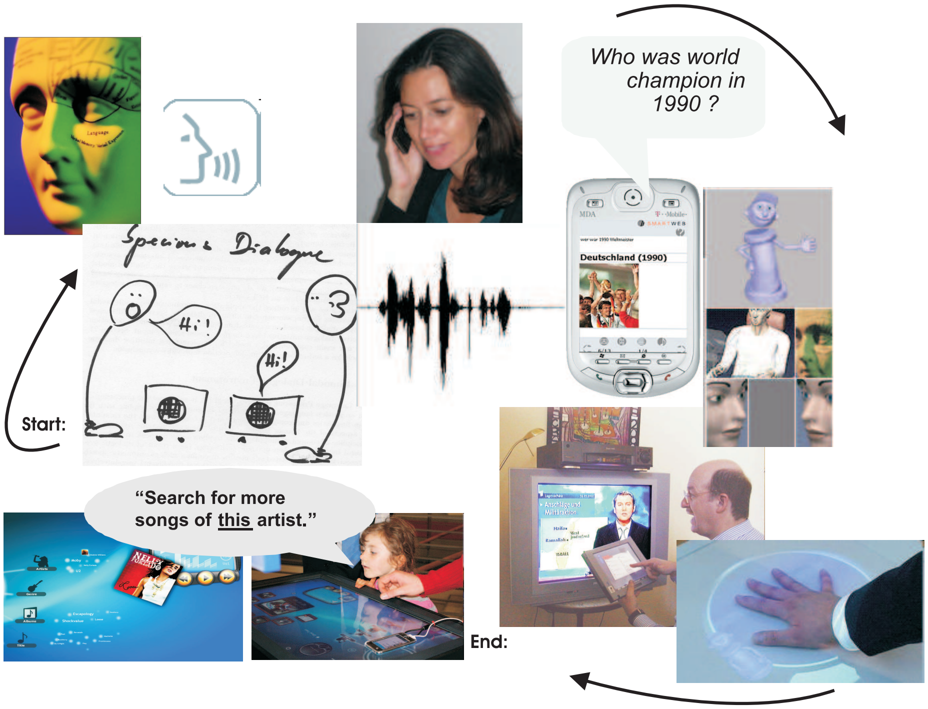

Who was world champion in 1990 ?
Speech Dialogue
Hi!
Hi!
Deutschland (1990)
Start:
“Search for more songs of this artist.”
End:

Fig. 1. Applications of spoken dialogue technology towards multimodal dialogue in mobile or ambient intelligence environments

Airline Travel Planning [22].[5] Speech-based question answering (QA) on mobile telephone devices is a natural extension of telephone-based dialogue systems [23]. Additionally, new smartphones should be able to answer questions about specific domains (e.g., the football domain) in real-time. User: “Who was world champion in 1990?”—System: “Germany.” Other directions (figure 1, down right) are the use of virtual human-like characters and multimodal, ambient intelligence environments where cameras detect human hand and finger gestures [24]. With the recent advent of more powerful mobile devices and APIs (e.g., the iPhone) it is possible to combine the dialogue system on the mobile device with a touchscreen table interaction. Multiple users can organise their information/knowledge space and share information with others, e.g., music files. These interaction modes combine Internet terminals and touchscreen access with mobile physical storage artefacts, i.e., the smartphones, and allow a casual user to search for and exchange music in an *intuitive* way. The role of intuition will be rendered more precisely in the rest of this chapter.

[5] The German Competence Center for Language Technology maintains a list of international research projects and available software in this area. Collate is one of these projects (http://collate.dfki.de/index_en.html).

In general, the interaction between the user and the multimodal dialogue system can either be user-initiative, system-initiative, or mixed-initiative. In user-initiative systems, the user has to utter a command before the system starts processing. Hence, the focus is more or less on the correct interpretation of the user's utterance; the new user intention, therefore, drives the behaviour of the dialogue system. A typical example can be found in route-planning in the transportation domain. For example, User: "I need to get to London in the afternoon"—System: "Ok. Query understood." The system can react by taking initiative and gathering the necessary information for the decision which mode of transportation would be preferable. In system-initiative systems, the user and the system must have a collaborative goal, and after initialisation (the user question), the system basically asks for missing information to achieve this goal. In [25], mixed-initiative is defined as: "[...] the phrase to refer broadly to methods that explicitly support an efficient, natural interleaving of contributions by users and automated services aimed at converging on solutions to problems." This basically means that the sub-goals and commitments have to come from both parties and be cleverly fulfilled and/or negotiated.

Why is this distinction in dialogue initiative so important to us for a better understanding of intuitive multimodal dialogue systems?

First, we need to interpret others' actions properly in order to respond or react in an expected and collaborative way. This also means that we must be able to interpret the input consistently in accordance with one's beliefs and desires. There are active and passive modes of conversation. Especially the passive modes are often unintentional, but we can perceive them *instinctively.* In the first instance, the dialogue system's perception of the *passive* input modes, e.g., a gaze through multimodal sensories, allows it to maintain a model of instinctive dialogue initiative as system-initiative.

Second, it is important to point out that the user/system initiative is not only about natural language. Multiple input and output signals in different modalities can be used to convey important information about (common) goals, beliefs, and intentions. In most of our examples, dialogue is in service of collaboration. However, the goal is still to solve a specific problem. Dialogue initiative, and mixed-initiative interaction, arises naturally from the joint intentions and intuitions about how to best address the problem solving task, thereby forming a theory of collaborative activity. The following human-human dialogue illustrates the introduction of a dialogue topic, a topic shift, and a development of a common interest that is pursued as the goal of the conversation.

1. **A:** "Hey, so are you going to see that new movie with Michael Jackson?" (topic introduction)
2. **B:** "You mean 'This Is It'?"
3. **A:** "Yeah, I think that's what it's called."
4. **B:** "When is it coming out?"
5. **A:** "It should be out in about a week. I really hope they also include some of the songs from the Jackson Five years. Do you like them, too?" (topic shift)
6. **B:** "I really like the Jackson Five! It's too bad the solo albums from Jermaine Jackson never became that popular."
7. **A:** "Exactly! Did any of the other members of the group produce solo albums?" (development of common interest)

Interaction technology and HCIs are extremely popular. However, the technology is still in a stage of infancy, especially when it comes to the dialogue management task and the modelling of interaction behaviour as demonstrated in the dialogue example. Especially the development of a common interest demands for a fine-gained dialogue state model and intuitive capabilities to perceive the common interest. HCIs are also popular in the context of (effective) information retrieval. Advanced user interfaces, such as multimodal dialogue interfaces, should provide new solutions. The Information Retrieval (IR) application domain [26, 27] is very suitable for making a list of current challenges of multimodal dialogue for which a model of intuition should help to properly address these challenges. In the context of information retrieval interaction (also cf. [28]), we often work with systems that:

- cannot recognise and develop common goals (as demonstrated in the dialogue example)
- use canned dialogue segments (e.g., "The answer to your query [input x] is [input y].");
- use hardwired interaction sequences (no sub-dialogues or clarifications possible);
- do not use inference services (new information cannot be used to infer new knowledge); and
- have very limited adaptation possibilities (e.g., no context adaptation is possible).

Combined methods to overcome the limitations include: (1) obeying dialogue constraints, (2) using sensory methods, and (3) modelling self-reflection and adaptation to implementing intuition (chapter 5). We can work against the limitations of current multimedia and HCI technology by exploiting dialogue systems with special (metacognitive) abilities and interaction agents that can simulate instincts and use intuitive models. We distinguish foraging, vigilance, reproduction, intuition, and learning as the human basic instincts (also cf. [29]). However, foraging and reproduction have no embodiment in contemporary AI for interaction technology and HCIs.

Figure 2 shows the dependencies and influences of adaptivity and instincts of a dialogue environment towards implementing intuition in dialogue. Adaptive dialogue might be the first step to overcome the difficulties. A context model influences the adaptive dialogue possibilities. Adaptive dialogue is a precondition for implementing instinctive dialogue since instincts are the triggers to adapt, e.g., the system-initiative. Multimodal sensory input influences instinctive dialogue. Instinctive dialogue is necessary for implementing intuitive dialogue. Instincts provide the input for a higher-order reasoning and intuition model. A self-reflective model/machine learning (ML) model influences intuitive dialogue. The strong relationship between *instinctive computing* and *multimodal dialogue systems* should enable us to introduce the notion of *intuition* into multimodal dialogue. Implementing intuition in dialogue is our main goal and, as we will explain in the rest of this chapter, instinctive dialogue in the form of an instinctive dialogue initiative is seen as the precondition for implementing intuition.

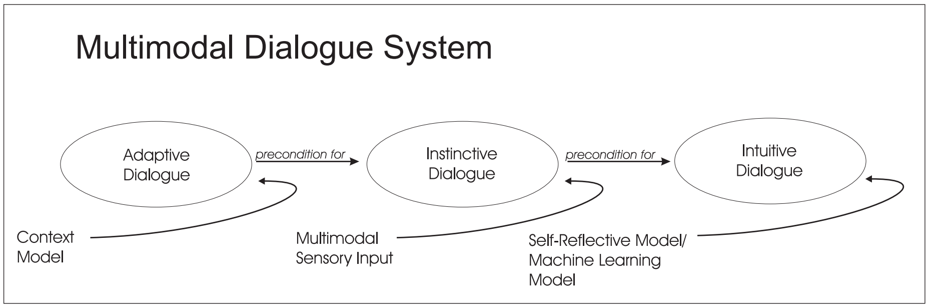


Fig. 2. Multimodal dialogue system properties: dependencies and influences of adaptivity and instincts of a dialogue environment towards implementing intuition in dialogue

3 Adaptive Dialogue

Adaptive dialogue systems can handle errors that occur during dialogue processing. For example, if the ASR recognition rate is low (i.e., the user utterances cannot be understood), an adaptive system can proactively change from user/mixed-initiative to system-initiative and ask form-filling questions where the user only responds with single open-domain words like surnames, “Yes”, or “No”. Adaptive dialogue systems allow these changes in dialogue strategies not only based on the progression of the dialogue, but also based on a specific user model. Additionally, specific dialogue obligations constrain the search space of suitable dialogue strategies [30–34]. User modelling and dialogue constraints will be explained in more detail in the following two subsections.

3.1 User Modelling

Incorporation of user models (cf., e.g., a technical user modelling architecture in [35]) helps novice users to complete system interaction more successfully and quickly as much help information is presented by the system. Expert users, however, do not need this much help to perform daily work tasks. As help information and explicit instructions and confirmations given by the system increase the number of system turns, communications get less efficient. Only recently has user modeling based performance analysis been investigated. For example, [36] tried to empirically provide a basis for future investigations into whether adaptive system performance can improve by adapting to user uncertainty differently based on the user class, i.e., the user model. Likewise, [37] argue that NLP systems consult user models in order to improve their understanding of users’ requirements and to generate appropriate and relevant responses. However, humans often *instinctively* know when their dialogue contribution is appropriate (e.g., when they should speak in a formal meeting or reduce the length of their contribution), or what kind of remark would be relevant in a specific dialogue situation.

3.2 Dialogue Constraints and Obligations

Dialogue constraints subsume four constraint types: linguistic constraints (e.g., correct case and number generation), correct dialogue acts as system responses (cf. adjacency pairs, answers should follow user questions, for example), timing constraints, and constraints on the information content itself (e.g., information to be presented should generally follow Grice's maxims and the users' presumptions about utterances; information should be available in an appropriate quantity).

In a dialogue, participants are governed by beliefs, desires, intentions (BDI), but also obligations. Beliefs represent the current state of information and are what a dialogue participant believes in terms of the other dialogue participant(s) and her world in general. Desires are what the dialogue participant would like to accomplish; they represent her source of motivation and are rather general. It is possible to have two or more desires which are not possible simultaneously (for example, wanting to go to Bombay and Paris on the next vacation). Intentions describe what the dialogue participant has chosen to do. At this point, the dialogue participant has already decided on a desire to pursue.

Obligations are the results of (largely social) rules or norms by which a dialogue participant lives. They play a central role when understanding and creating dialogue acts for a natural dialogue because they account for many of the next user or system moves during the dialogical interaction. Usually, a general cooperation between dialogue participants is assumed based on their intentions and common goals. However, this assumption fails to acknowledge a not infrequently uncooperativeness in respect to shared conversational goals. Recognising this dialogue behaviour in other dialogue participants can help develop more effective dialogues toward *intuition in dialogue* by changing the dialogue strategy.

Obligations are induced by a set of social conventions. For instance, when a dialogue participant asks a question, the obligation is to respond. It is therefore relevant how obligations can be identified, which rules can be developed out of this knowledge, and how the system can properly respond. For example, when a participant uses the interjection "uh" three times, the system responds by helping the user, or when a participant speaks, the system does not interrupt. We can say that the system adapts to the user model while at the same time obeying the dialogue obligations. (Also see [38] for a list of social discourse obligations.) The following lists give examples for important social obligations we encounter in terms of instinctive and intuitive dialogue when assuming that the conversational goal supports the task of the user (also cf. [34], page 149ff.). The core social discourse obligations are:

1. self-introduction and salutation: *"Hi there." "Hello." "Good morning/evening." "Hi. How can I help you?" "What can I do for you?" "Hi [name], how are you today?"*
2. apology: *"I'm so sorry." "I'm sorry." "It seems I've made an error. I apologise."*
3. gratitude: *"Thank you so much." "I appreciate it." "Thank you."*
4. stalling and pausing: *"Give me a moment, please." "One minute." "Hang on a second."*

In the context of information-seeking multimodal dialogue, extra-linguistic inputs/outputs (i.e., anything in the world outside language and the language modality, but which is relevant to the multimodal utterance) and social obligations have particular multimodal implementations:[6]

1. Maintaining the user's attention by reporting on the question processing status (+ special gestures in case of embodied agents, e.g., looking into the others' eyes; + special mimics). If a process is successful, e.g., embodied agents can express joy. If the query cannot be processed satifactorily, the linguistic output can be accompanied with expressions of shame (figure 3):
   - *"Well, it looks like this may take a few minutes."*
   - *"I'm almost done finding the answer to your question."*
   - *"Give me a moment and I'll try to come up with something for you."*
   - *"I'll have your answer ready in a minute."*
   - *"I am getting a lot of results so be patient with me."*
2. Informing the user about the probability of query success, i.e., the probability the user is presented the desired information, or informing the user as to why the current answering process is due to fail (+ special gestures such as shaking one's head or to shrugging one's shoulders):
   - *"There is a good chance that I will be able to come up with a number of possibilities/solutions/answers for you."*
   - *"I will probably get a lot of results. Maybe we can narrow them down already?"*
   - *"Sorry, but apparently there are no results for this enquiry. I can try something else if you like."*
   - *"This search is taking longer than it should. I probably won't be able to come up with an answer."*
   - *"It seems that I can only give you an answer if I use a special service/program. This is not a free service, though."*
3. Balancing the user and system initiative (equipollent partners should have the same proportion of speech in which they can point something out).

4 Instinctive Dialogue

We hypothesise that underlying maxims of conversation and the resulting multimodal dialogue constraints may very much be related to instinctive computing [29]. The instincts discussed here are vigilance, learning, and intuition, whereby intuition is considered a cognitively more advanced form of intelligence which builds on the input of the other instincts, vigilance and learning. Human-centred human-computer interaction strategies are applied to enable computers to unobtrusively respond to the user-perceived content. These strategies can be based on instinctive reactions which take vigilance and learning into account.

[6] Extra-lingustic universals in communication and language use are also described in the context of politeness, see [39].

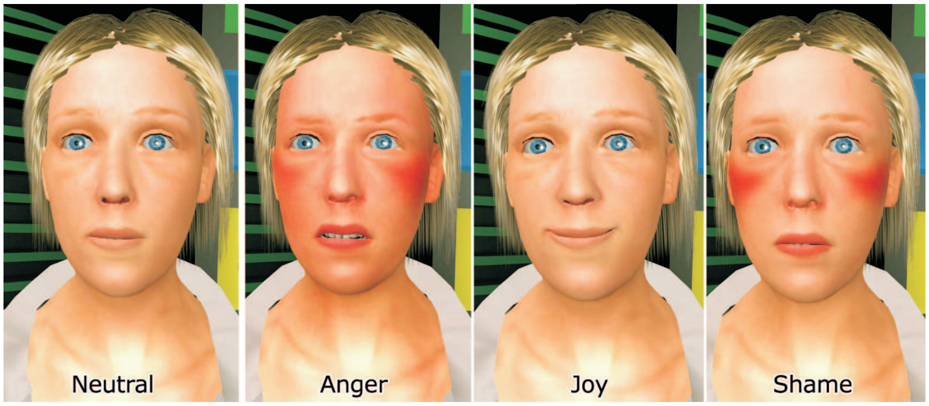


Fig. 3. Emotional expressions in embodied virtual agents. These examples are taken from the VirtualHumans project, see http://www.virtual-human.org/ for more information.

We attempt to shed light on the relationship between instinctive computing and state-of-the-art multimodal dialogue systems in order to overcome the limitations of contemporary HCI technology. Instincts relate to dialogue constraints (as explained in chapter 3), and the attempts to convey them, in order to make HCIs more intelligent. Linguistic features (syntax, morphology, phonetics, graphemics), para-linguistic features (tone, volume, pitch, speed, affective aspects), and extra-linguistic features (haptics, proxemics[7], kinesics, olfactics, chronemics) can be used to model the system state and the user state (e.g., emotional state or stress level). Multimodal sensory input recognition, and the recognised affective states (e.g., laughing) and emotions, play the central roles in instinctive dialogue.

4.1 Multimodal Sensory Input

Sensors convert a physical signal (e.g., spoken language, gaze) to an electrical one that can be manipulated symbolically within a computer. This means we interpret the spoken language by ASR into text symbols, or a specific gaze expression into an ontology-based description, e.g., the moods exuberant and bored. An ontology is a specification of a conceptualisation [40] and provides the symbols/names for the input states we try to distinguish.

As mentioned before, there are passive and active sensory inputs. The passive input modes, such as anxiety or indulgence in a facial expression, roughly correspond to the ones perceived instinctively. (This is also consistent with our definition of intuition since many passive sensory input modes are not consciously perceived by the user.) The active input modes, on the other hand, are mostly

[7] "The branch of knowledge that deals with the amount of space that people feel it necessary to set between themselves and others." (New Oxford Dictionary of English)

the linguistic features such as language syntax and morphology. These convey the proposition of a sentence (in speech theory this means the content of a sentence, i.e., what it expresses in the specific context). The passive input modes are, however, more relevant for modelling users' natural multimodal communication patterns, e.g., variation in speech and pen pressure [33]. That is, users often engage in hyperarticulate speech with computers (as they would be talking to a deaf person). Because they expect computers to be error-prone, durational and articulatory effects can be detected by ASR components (also cf. [41]).

Multimodal interaction scenarios and user interfaces may comprise of many different sensory inputs. For example, speech can be recorded by a bluetooth microphone and sent to an automatic speech recogniser; camera signals can be used to capture facial expressions; the user state can be extracted using biosignal input, in order to interpret the user's current stress level (e.g., detectors measuring the levels of perspiration). Stress level corresponds to an instinctive preliminary estimate of a dialogue participant's emotional state (e.g., anger vs. joy). In addition, several other sensory methods can be used to determine a dialogue's situational and discourse context—all of which can be seen as an instinctive sensory input. First, the attention detection detects the current focus of the user by using on-view/off-view classifiers. If you are addressed with the eyes in, e.g., a multi-party conversation, you are more vigilant and aware that you will be the next to take over the dialogue initiative. Therefore you listen more closely; a computer may activate the ASR (figure 4.1, right).

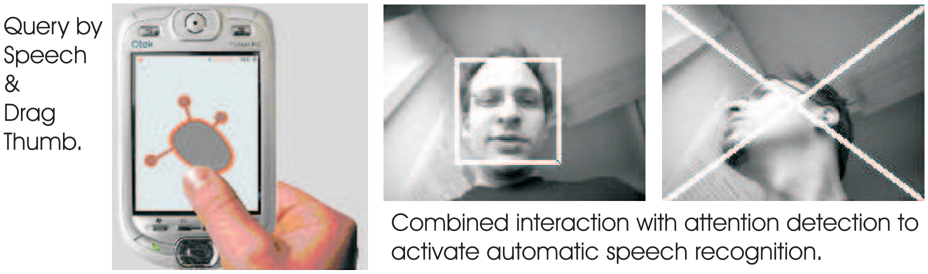

Query by Speech & Drag Thumb.
Combined interaction with attention detection to activate automatic speech recognition.

Fig. 4. Anthropocentric thumb sensory input on mobile touchscreen (left) and two still images illustrating the function of the on-view/off-view classifier (right).

Passive sensory input, e.g., gaze, still has to be adequately modelled. People frequently direct their gaze at a computer when talking to a human peer. On the other hand, while using mobile devices, users can be trained to direct their gaze toward the interaction device (e.g., by providing direct feedback for the user utterance). This can enhance the usability of an interaction device enormously when using an open-microphone engagement technique with gaze direction input. Although good results have been obtained for the continuous listening for unconstrained spoken dialogue [42], passive multimodal sensory inputs offer additional triggers to direct speech recognition and interpretation.

In addition, people have limited awareness of the changes they make when addressing different interlocutors, e.g., changes in amplitude are not actively perceived. On the other hand, the addressee of the message may be very sensitive to changes in amplitude he perceives. For example, think of how some people's voices change when they speak to small children or their spouses. ASR components can be trimmed to emphasis detection. An important theoretical framework that the research in, e.g., [33] builds on is that the hypo-to-hyper emphasis spectrum is characteristic not only for speech (e.g., hyper-articulation and hyper-amplitude), but for all modes of communication. In particular, a test can be constructed around whether an utterance was intended as a request to the computer. The result would be an instinctive addressee sensor the multimodal dialogue system could use as passive trigger for, e.g., system initiative.

4.2 Recognising Moods and Emotions

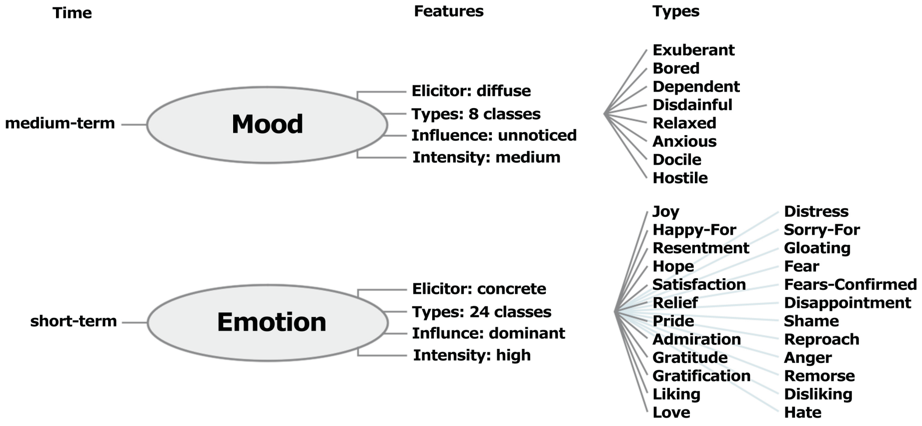


Fig. 5. Moods and emotions according to [43]

Feelings and emotions have been discussed in relevant psychological literature, e.g., [44]. More modern, and technically grounded, works speak of moods and emotions in the mind, e.g., [45]. We use the distinction between moods and emotions (figure 5). The realisation of emotions (in embodied virtual characters, such as the ones in figure 8, right) for speech and body graphics has been studied, e.g., in [46]. [43] used more fine-grained rules to realise emotions by mimicry, gesture, and face texture. Moods are realised as body animations of posture and gestures. Affective dialogue systems have been described in [47].[8]

[8] Important relates work comprises of [48] who outlines a general emotion-based theory of temperament. [49] deals with the cognitive structure of emotions, whereby [50] introduces a five-factor model of personality and its applications.

Computational models for emotions are explained in [51] and bring together common sense thinking and AI methods. In multimodal dialogue, [52] discusses emotion-sensing in life-like communication agents and [53] describes the history of the automatic recognition of emotions in speech while presenting acoustic and linguistic features used. There are many interesting new projects for this purpose, too: *Prosody for Dialog Systems*[9] investigates the use of prosody, i.e., the rhythm and melody of speech in voice input, to human-computer dialog systems. Another project, HUMAINE[10], aims to develop systems that can register, model, and/or influence human emotional and emotion-related states and processes (works towards emotion-oriented interaction systems). EMBOTS[11] create realistic animations of nonverbal behaviour such as gesture, gaze, and posture.

In the context of corpus-based speech analysis, [54] find a strong dependency between recognition problems in the previous turn (a turn is what a dialogue partner says until another dialogue partner rises to speak) and user emotion in the current turn; after a system rejection there are more emotional user turns than expected. [55] address a topic and scenario that we will use to formulate intuitive dialogue responses in an example dialogue, i.e., the modelling of student uncertainty in order to improve performance metrics including student learning, persistence, and system usability. One aspect of intuitive dialogue-based interfaces is that one of the goals is to tailor them to individual users. The dialogue system adapts (intuitively) to a specific user model according to the sensory input it gets.

4.3 Towards Intuition

When it comes to implementing intuition, we expect an *instinctive interaction agent* to deliver the appropriate sensory input. A useful and cooperative dialogue in natural language would not only combine different topics, heterogeneous information sources, and user feedback, but also intuitive (meta) dialogue—initiated by the instinctive interaction agent. Many competences for obeying dialogue constraints fall into the gray zone between competences that derive from instincts or intuition. The following list enumerates some related aspects:

- As mentioned before, instinctive and intuitive dialogue interfaces should tailor to individual users. The dialogue system adapts *intuitively* to a specific user model according to the sensory input it gets *instinctively.*
- Different addressees can often be separated intuitively, not only by enhancing the intelligibility of the spoken utterances, but by identifying an intended addressee (e.g., to increase amplitude for distant interlocutors). In this context, gaze represents a pivotal sensory input. [56] describes an improved classification of gaze behaviour relative to the simple classifier "the addressee is where the eye is."

[9] http://www.speech.sri.com/projects/dialog-prosody

[10] http://emotion-research.net

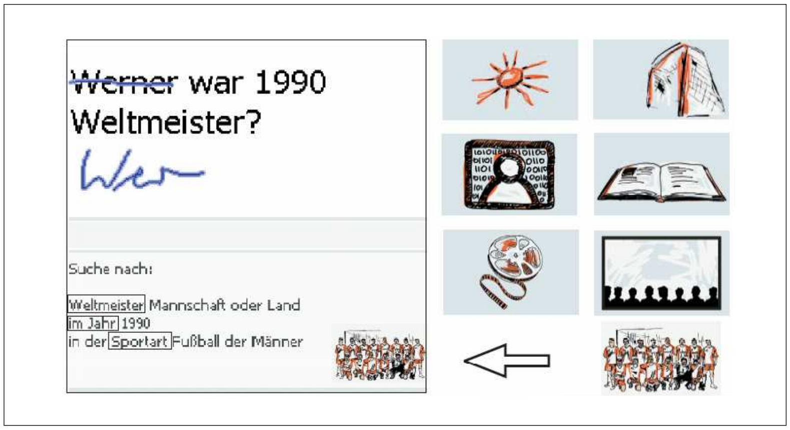
[11] http://embots.dfki.de

- Intuitive dialogue means using implicit user communication cues (meaning no explicit instruction has come from the user). In this case, intuition can be defined as an estimation of a user-centred threshold for detecting when the system is addressed in order to (1) automatically engage, (2) process the request that follows the ASR output in the dialogue system, and (3) finally respond.
- With anthropocentric interaction design and models, we seek to build input devices that can be used intuitively. We recognised that the thumb plays a significant role in modern society, becoming humans' dominant haptic interactor, i.e., main mode of haptic input. This development should be reflected in the interface design for future HCIs. Whether society-based interaction habits (e.g., you subconsciously decide to press a doorbell with your thumb) can be called an instinctive way of interacting, is just one aspect of the debate about the relationship between intuition and instincts (figure 4.1, left). The combination of thumb sensory input and on-view/off-view recognition to trigger ASR activation is very intuitive for the user and reflects instinctive capabilities of the dialogue system toward intuition capabilities.
- Intuitive question feedback (figure 6), i.e., graphical layout implementation for user queries, is a final aspect (generation aspect) of instinctive and intuitive dialogue-based interfaces. The question *"Who was world champion in 1990?"* results in the augmented paraphrase *Search for: World champion team or country in the year 1990 in the sport of football, division men.* Concept icons (i.e., icons for the concepts cinema, book, goal, or football match, etc.) present feedback demonstrating question understanding (a team instance is being asked for) and answer presentation in a language-independent, intuitive way. The sun icon, for example, additionally complements a textual weather forecast result and conveys weather condition information.

Intuition can be seen as instinctive dialogue. Intuition can also be seen as cognitively advanced instincts.

5 Intuitive Dialogue

The definition of intuition we introduced—the power of obtaining knowledge that cannot be acquired either by inference or observation, by reason or experience—is unsustainable in the context of technical AI systems. The previous sections have given examples of intuitive behaviour only made possible by inference or observation. Instead we say that intuition is based on inference or observation, by reason or experience, but the process happens unconsciously in humans. This gives us the freedom to use technical (meta) cognitive architectures towards technical models of intuition, based on perception of moods and emotions in addition to the language in a multimodal dialogue system. Mood recognition has mostly been studied in the context of gestures and postures (see, e.g., [57]). Boredom has visual gesture indicators, e.g., looking at the watch, yawning (also



~~Werner~~ war 1990 Weltmeister?

Wer

Suche nach:

Weltmeister Mannschaft oder Land
im Jahr 1990
in der Sportart Fußball der Männer

Fig. 6. Intuitive question feedback

cf. “a bit of a yawn”, and putting the hand on the mouth), posture buckled upper part of the body, slouchy, slow basic motion.

Mood recognition has mostly been studied in the context of gestures and postures (see, e.g., [57]). Boredom has visual gesture indicators, e.g., looking at the watch, yawning (also cf. “a bit of a yawn”, and putting the hand to the mouth), the posture of a buckled upper part of the body, and slouchy or slow basic motions. However, the mimicry of boredom is much more difficult to describe as, e.g., the mimicry of the basic emotions (figure 3). This also means that automatic facial expression methods (e.g., [58], who use robust facial expression recognition from face video) have difficulties in detecting this even when analysing the temporal behaviour of the facial muscle units. In [59], 3D wireframe face models were used to discriminate happiness from anger and occluded faces. However, more complex expressions as in figure 7 cannot be detected with the required accuracy.[12] The recognition on a 2D surface is even more difficult. How comes that humans are able to easily detect eagerness and boredom in watercoloured coal drawings? Especially the correct interpretation of these non-verbal behavioural signals is paramount for an intuitive reaction in multimodal dialogue, e.g., a proposition for personal recreational activities (based on context information and a user profile, see figure 8, left). The following dialogue example (adapted and extended from [46]) exemplifies intuitive reaction behaviour of the host (H) as a result of his intuitive perception of the emotions and moods of the players Mr. Kaiser and Ms. Scherer in the football game (figure 8, right).

1. **H:** “Now, pay attention [points to paused video behind them] What will happen next? One—Ballack will score, Two—the goalie makes a save, or Three—Ballack misses the shot?”
2. **H:** “What do you think, Mr. Kaiser?”

[12] See more facial expressions and emotions, and other non-verbal (extra-linguistic) behavioural signals in [60].

3. **K:** “I think Ballack’s going to score!”[13]
4. **H:** “Spoken like a true soccer coach.”[14]
5. **H:** “Then let’s take a look and see what happens!”
   (All turn and watch the screen. Ballack shoots but misses the goal.)
6. **H:** “Well, it seems you were really wrong!”[15]
7. **H:** “But, don’t worry Mr. Kaiser, this isn’t over yet. Remember, you still have two more chances to guess correctly!”[16]
8. **H:** “This time, um, I guess the goalie will make a save.”[17]
9. **H:** “Ms. Scherer, do you agree with Mr. Kaiser this time?” (active violation of the one player-host question-answer pattern)
10. **S:** “Oh, uh, I guess so.”[18]
11. **H:** “Great! Well, then let’s see what happens this time!”[19]

Intuition in dialogue systems also has a strong active component. It should help the system to react in an appropriate way, e.g., to avoid distracting the user such that cognitive load remains low and the user can still focus on the primary task, or, to motivate, convince, or persuade the user to develop an idea further and/or pursue mutual goals.

Techniques for information fusion are at the heart of intuitive dialogue design. Humans are capable of correctly fusing different information streams. For example, think of the combination of spoken language and visual scene analysis you have to perform when a passenger explains the route to you while you are driving. People can automatically adapt to an interlocutor’s dominant integration pattern (parallel or sequential). This means, e.g., while explaining the route to you, your friend points to a street and says “This way!”; he can do this in a parallel or sequential output integration pattern and you adapt accordingly. Whereas the on-focus/off-focus detection alone (figure 4.1, right) cannot really be seen as intuitive, the combination of automatically-triggered speech recognition with deictic reference fusion (“Is this (+ deictic pointing gesture) really the correct way?”) and gesture recognition for dissatisfaction/unhappiness actually can be seen as intuitive. The system can perceive dissatisfaction and suggest a query reformulation step or other query processing adaptions.

5.1 Self-Reflective Model

We believe that a self-reflective model is what accounts for intuitive behaviour. Psychological literature provides the necessary concept of such a self-reflective

[13] Mr. Kaiser interprets the host’s question as something positive, since he was asked and not Ms. Scherer.

[14] The host intuitively perceives the certainty in Mr. Kaiser’s reaction. According to [61] (their related context is tutoring dialogues for students) the host perceives a InonU (incorrect and not uncertain) or CnonU (correct and not uncertain) reaction.

[15] The host recognises the InonU case. Mr. Kaiser looks down, is very disappointed (and knows he should not have been that cheeky). Ms. Scherer looks pleased.

[16] The host recognises that Mr. Kaiser is disappointed and that Ms. Scherer is mischievous.

[17] Host notices that Mr. Kaiser is now more moderate and Ms. Scherer is not paying attention anymore and tries to include her as well.

[18] Ms. Scherer realises she has not been paying attention and is embarrassed about having been caught. Immediately, she is also relieved, though, since she realises that the host is not trying to make this evident but bring her focus back to the game.

[19] The host knows the focus of both players is back and goes on with the game.

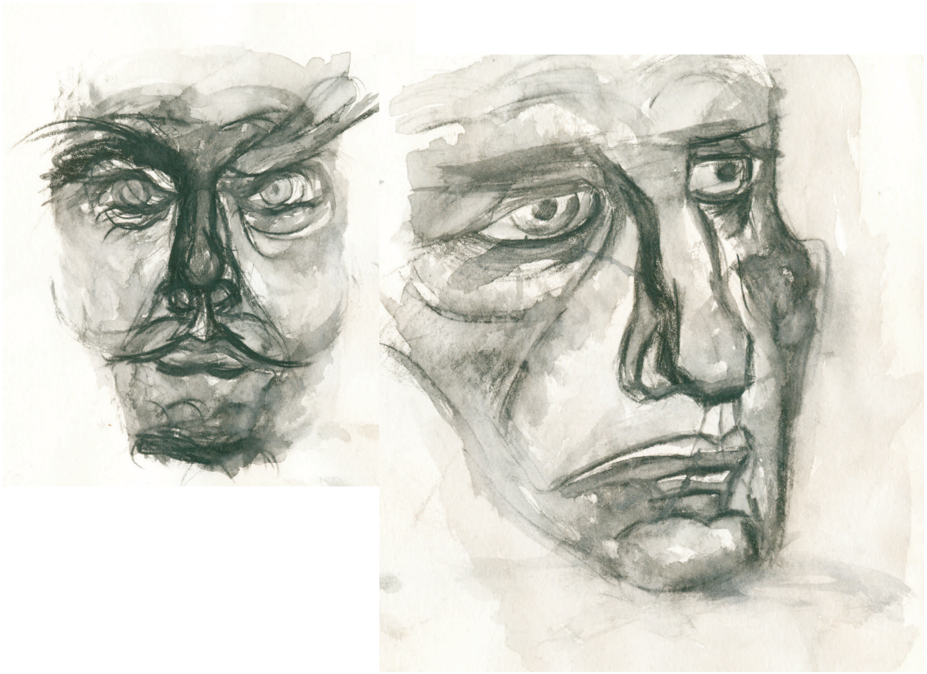


Fig. 7. (Left) Eager demand for a reaction. (Right) No resources against boredom.

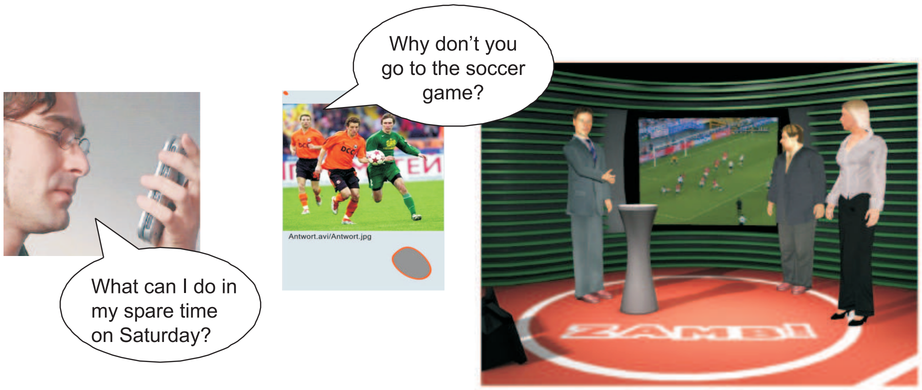

Why don't you go to the soccer game?
What can I do in my spare time on Saturday?
Antwort.avi/Antwort.jpg
ZAMBI

Fig. 8. (Left) Intuitive soccer dialogue through context information with the help of a mobile dialogue system. (Right) Dialogue with virtual characters in a virtual studio (reprinted with permission [46]). In the mobile situation, intuition is needed to understand the intentions of the user (perception). In the studio, intuition can lead to emotion expression and emphatic group behaviour (generation).

model. Therefore, we will introduce self-reflection by applying the general processing framework of metacognition by [62]. In their theoretical metamemory framework, they base their analysis of metacognition on three principles:

1. *The cognitive processes are split into two or more interrelated levels.*
2. *The meta-level (metacognition) contains a dynamic model of the object-level (cognition).*
3. *The two dominant relations between the levels are called* **control** *and* **monitoring**.

The basic conceptual architecture consists of two levels, the object-level and the meta-level (figure 9). We will use the two level structure to model a self-reflective model for implementing intuition. (Specialisations to more than two levels have also been developed, see [63]). This architecture extends approaches to multi-strategy dialogue management, see [64] for example.

The main point is that we can maintain a model of the dialogue environment on the meta-level which contains the context model, gets access to the multimodal sensory input, and also contains self-reflective information. We called such a model an *introspective view*, see figure 10. An introspective view emerges from monitoring the object level—the correct interpretation of available sensory inputs and the combination of this information with prior knowledge and experiences. We can use this knowledge to implement a dialogue reaction behaviour (cf. [34], p. 194f) that can be called intuitive since the object level control fulfills intuitive functions, i.e., the initiation, maintenance, or termination of object-level cognitive activities; our intuition controls our object-level behaviour by formulating dialogue goals and triggering dialogue actions.

Machine Learning Model What is worth learning in the context of intuition? For example learning about the adaptation to differences in user integration patterns (i.e., not to wait for multimodal input when the input is unimodal would be beneficial). The understanding of the user's (temporal) integration patterns of multiple input modalities can be seen as intuitively understanding how to fuse the passive sensory input modes of a user. This can be seen as a machine learning classification task. The problem identification step (diagnosis) will be supported by data mining models which are learned by mining the process data log files which will have been obtained by running the baseline system.

Unsupervised association rule generation can be employed to identify rules of failure and success according to the item sets derived from processing metadata.[20]

Methodology to Implement Intuition We will briefly discuss a new methodology of system improvements to ensure future performances for implementing intuition in dialogue systems. The methodology and methods for self-reflection in dialogue systems consist of the following data and model resources (delivering ML input data or ML models):

[20] For more information, see [34], pp. 179-208.

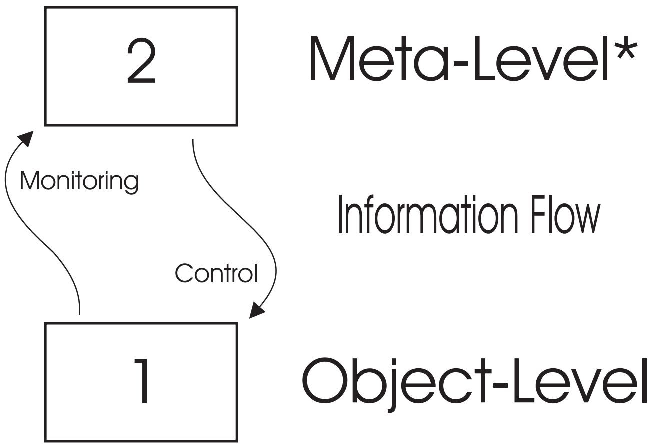
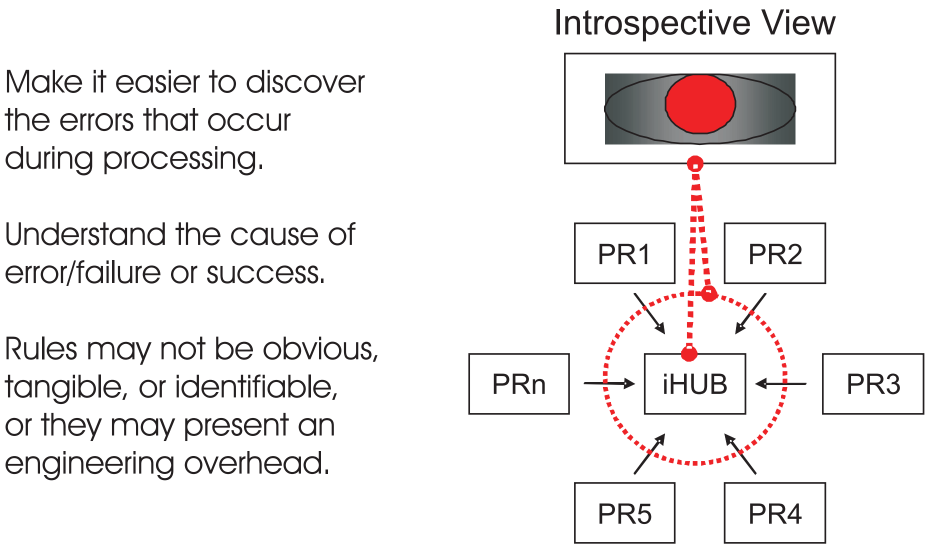


Fig. 9. The self-reflective mechanism consists of two structures, (1) the object-level, and (2) the meta-level, whereby 1 → 2 is an asymmetric relation *monitoring*, and 2 → 1 is an asymmetric relation *control*. Both form the information flow between the two levels; monitoring informs the meta-level and allows the meta-level to be updated. Depending on the meta-level, the object-level is controlled, i.e., to initiate, maintain, or terminate object-level cognitive activities like information retrieval or other dialogue actions.


Make it easier to discover the errors that occur during processing.

Understand the cause of error/failure or success.

Rules may not be obvious, tangible, or identifiable, or they may present an engineering overhead.

Fig. 10. Introspective View. Several processing resources (PRs), such as sensories, are connected to a central dialogue system hub. The iHUB interprets PR inputs and outputs on a meta level.

- *Cognitive Ground*: The theory assumes a baseline functional dialogue reaction and presentation manager. Such a manager has been developed in [34]. Possible system improvements are identified by reaction utility shortcomings that become obvious by running the baseline finite state-based dialogue system in dialogue sessions with real users. According to this empirical dialogue system evaluation, a subset of the theoretical reaction constraints and utilities (cf. dialogue constraints implementing intuition) can be identified by the dialogue system experts and is improved by the supplied ML methods.
- *Information States*: Information states will be implemented by a local and a global state. Global states hold information about the current dialogue session and user profiles/models. The local state contains information about the current turn and the changes in the system or user model.
- *Feature Extraction*: In the feature extraction phase we extract features for the different empirical machine learning models. Algorithm performances on different feature subsets provide *a posteriori* evidence for the usefulness of individual features.
- *Associations*: Domain knowledge is often declarative, whereas control knowledge is more operational, which means that it can change over time or can only be correctly modelled *a posteriori.* Associations bridge the gap between cognition towards metacognition and intuitions. The cognitive ground can be used to induce dynamic associations for adaptation purposes.

Implemented Dialogue Feedback Example Intuitive question feedback by using concept icons has been shown in figure 6. This example showed a multimodal speech-based question answering (QA) dialogue on mobile telephone devices [23]. When human users intuitively adapt to their dialogue partners, they try to make the conversation informative and relevant. We can also assume that they avoid saying falsehoods or that they indicate a lack of adequate evidence. In order to learn similar intuitive dialogical interaction capabilities for question answering applications, we used the aforementioned methodology to implement intuition in information-seeking dialogues. In the information and knowledge retrieval context, information sources may change their quality characteristics, e.g., accessibility, response time, and reliability. Therefore, we implemented an introspective view on the processing workflow: machine learning methods update the reasoning process for dialogue decisions in order to allow the dialogue system to provide intuitive dialogue feedback. More precisely, we tried to incorporate intuitive meta dialogue when interpreting the user question and addressing heterogeneous information sources.

For example, we ran a baseline system, recorded the current state of the dialogue system, extracted features according to the introspective view/sensory input, and tried to generalise the found associations to knowledge that allows for more intuitive system reactions [34].

The associations we extracted basically revealed which types of questions we are able to answer with the current databases, how long it might take to answer a specific request, and how reliable an answer from an open-domain search engine might be. We could say that the system has an intuition of the probability of success or failure. This intuition (perceived/learned model about the dialogue environment) can be used to provide more intuitive question feedback of the forms: “My intuition says that... I cannot find it in my knowledge base” (stream/answer time prediction); “... I should better search the Internet for a suitable answer” (database prediction); or “... empty results are not expected, but the results won’t be entirely certain.” (answer prediction). A combination of the last two predictive models even allows a dialogue system to intuitively rely on specific open domain QA results (cf. the dialogue fragment further down). The Answer type *Person* is predicted to be highly confident for open-domain QA, so the system provides a short answer to the Person question instead of a list of documents as answer.

1. **U:** “Who is the German Chancellor?”
2. **S:** “Who is the German Chancellor?” (intuitively repeats a difficult question)
3. **S:** “I will search the Internet for a suitable answer.”
4. **S:** “Angela Merkel.” (intuitively relies on the first entry in the result set)

Although the exemplified intuitive behaviour is by far less complicated, or intuitive, than the host’s behaviour in the envisioned example of the intuitive dialogue section, we think that the self-reflective model is an important step towards further achievements in the area of implementations of intuition in multimodal dialogue.

6 Conclusion

Intelligent user interfaces have to become more human-centred. This is not only because state-of-the-art HCIs are far beyond human interaction capabilities, but also because the amount of information to be processed is constantly growing. This makes the automatic selection of suitable information more complicated, and personalised and adapted user interfaces more valuable. [65] argue that services (e.g., for information retrieval) should be organised in a service-oriented architecture that enables self-organisation of ambient services in order to support the users’ activities and goals. User interfaces that instinctively take initiative during multimodal dialogue to achieve a common goal, which is negotiated with the user, provide new opportunities and research directions. Intuitive dialogue goes one step further. It does not only take the multimodal sensory input spaces into account to trigger instinctive dialogue reaction rules, but allows for maintaining a self-reflective model. This model can evolve over time with the help of learned action rules. If a model is updated and applied unconsciously, we can speak of modelling and implementing intuition in multimodal dialogue.

One of the major questions for further debates would be whether appropriate intelligent behaviour in special situations, such as intuition in natural multimodal dialogue situations, is more rooted in past experience (as argued here in the context of intuition and learning) than logical deduction or other relatives in planning and logical reasoning.

Acknowledgements This research has been supported in part by the THESEUS Programme in the Core Technology Cluster WP4, which is funded by the German Federal Ministry of Economics and Technology (01MQ07016). The responsibility for this publication lies with the author. I would like to thank Colette Weihrauch for the help in the formulation of dialogue constraints and related example dialogues. Thanks to Ingrid Zukerman for useful conversations.

References

1. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. 2nd edition edn. Prentice-Hall, Englewood Cliffs, NJ (2003)
2. Mitchell, T.M.: Machine Learning. McGraw-Hill International Edit (1997)
3. Maybury, M., Stock, O., Wahlster, W.: Intelligent interactive entertainment grand challenges. IEEE Intelligent Systems **21**(5) (2006) 14–18
4. Maybury, M.T. In: Planning multimedia explanations using communicative acts. American Association for Artificial Intelligence, Menlo Park, CA, USA (1993) 59–74
5. Sonntag, D.: Introspection and adaptable model integration for dialogue-based question answering. In: Proceedings of the Twenty-first International Joint Conferences on Artificial Intelligence (IJCAI). (2009)
6. Cole, R.A., Mariani, J., Uszkoreit, H., Varile, G., Zaenen, A., Zue, V., Zampolli, A., eds.: Survey of the State of the Art in Human Language Technology. Cambridge University Press and Giardini, New York, NY, USA (1997)
7. Jelinek, F.: Statistical Methods for Speech Recognition (Language, Speech, and Communication). The MIT Press (January 1998)
8. McTear, M.F.: Spoken dialogue technology: Enabling the conversational user interface. ACM Computing Survey **34**(1) (March 2002) 90–169
9. McTear, M.: Spoken Dialogue Technology. Springer, Berlin (2004)
10. Dybkjaer, L., Minker, W., eds.: Recent Trends in Discourse and Dialogue. Volume 39 of Text, Speech and Language Technology. Springer, Dordrecht (2008)
11. Maybury, M., Wahlster, W., eds.: Intelligent User Interfaces. Morgan Kaufmann, San Francisco (1998)
12. van Kuppevelt, J., Dybkjaer, L., Bernsen, N.O.: Advances in Natural Multimodal Dialogue Systems (Text, Speech and Language Technology). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
13. Woszczyna, M., Aoki-Waibel, N., Buo, F., Coccaro, N., Horiguchi, K., Kemp, T., Lavie, A., McNair, A., Polzin, T., Rogina, I., Rose, C., Schultz, T., Suhm, B., Tomita, M., Waibel, A.: Janus 93: towards spontaneous speech translation. Acoustics, Speech, and Signal Processing, IEEE International Conference on **1** (1994) 345–348
14. Wahlster, W., ed.: VERBMOBIL: Foundations of Speech-to-Speech Translation. Springer (2000)
15. Seneff, S., Hurley, E., Lau, R., Pao, C., Schmid, P., Zue, V.: Galaxy-II: A reference architecture for conversational system development. In: Proceedings of ICSLP-98. Volume 3. (1998) 931–934
16. Walker, M.A., Passonneau, R.J., Boland, J.E.: Quantitative and Qualitative Evaluation of the Darpa Communicator Spoken Dialogue Systems. In: Meeting of the Association for Computational Linguistics. (2001) 515–522

17. Walker, M.A., Rudnicky, A., Prasad, R., Aberdeen, J., Bratt, E.O., Garofolo, J., Hastie, H., Le, A., Pellom, B., Potamianos, A., Passonneau, R., Roukos, S., S, G., Seneff, S., Stallard, D.: Darpa communicator: Cross-system results for the 2001 evaluation. In: Proceedings of ICSLP. (2002) 269–272
18. Wahlster, W., ed.: SmartKom: Foundations of Multimodal Dialogue Systems. Springer, Berlin (2006)
19. Wahlster, W.: SmartWeb: Mobile Applications of the Semantic Web. In Dadam, P., Reichert, M., eds.: GI Jahrestagung 2004, Springer (2004) 26–27
20. Reithinger, N., Bergweiler, S., Engel, R., Herzog, G., Pfleger, N., Romanelli, M., Sonntag, D.: A Look Under the Hood—Design and Development of the First SmartWeb System Demonstrator. In: Proceedings of the 7th International Conference on Multimodal Interfaces (ICMI), Trento, Italy (2005)
21. Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N. [66] 272–295
22. Aaron, A., Chen, S., Cohen, P., Dharanipragada, S., Eide, E., Franz, M., Leroux, J.M., Luo, X., Maison, B., Mangu, L., Mathes, T., Novak, M., Olsen, P., Picheny, M., Printz, H., Ramabhadran, B., Sakrajda, A., Saon, G., Tydlitat, B., Visweswariah, K., Yuk, D.: Speech recognition for DARPA Communicator. Acoustics, Speech, and Signal Processing, IEEE International Conference on **1** (2001) 489–492
23. Sonntag, D., Reithinger, N.: SmartWeb Handheld Interaction: General Interactions and Result Display for User-System Multimodal Dialogue. Smartweb technical document (5), DFKI, Saarbruecken, Germany (2007)
24. Reithinger, N., Alexandersson, J., Becker, T., Blocher, A., Engel, R., Löckelt, M., Müller, J., Pfleger, N., Poller, P., Streit, M., Tschernomas, V.: SmartKom: Adaptive and Flexible Multimodal Access to Multiple Applications. In: Proceedings of the 5th Int. Conf. on Multimodal Interfaces, Vancouver, Canada, ACM Press (2003) 101–108
25. Horvitz, E.: Uncertainty, action, and interaction: In pursuit of mixed-initiative computing. IEEE Intelligent Systems **14** (1999) 17–20
26. van Rijsbergen, C.J.: Information Retrieval. 2 edn. Butterworths, London (1979)
27. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, New York, NY, USA (2008)
28. Ingwersen, P.: Information Retrieval Interaction. Taylor Graham, London (1992)
29. Cai, Y. [66] 17–46
30. Singh, S., Litman, D., Kearns, M., Walker, M.A.: Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System. Journal of Artificial Intelligence Research (JAIR) **16** (2002) 105–133
31. Jameson, A.: Adaptive interfaces and agents. In: The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA (2003) 305–330
32. Paek, T., Chickering, D.: The markov assumption in spoken dialogue management. In: Proceedings of the 6th SigDial Workshop on Discourse and Dialogue, Lisbon, Portugal (2005)
33. Oviatt, S., Swindells, C., Arthur, A.: Implicit user-adaptive system engagement in speech and pen interfaces. In: CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (2008) 969–978
34. Sonntag, D.: Ontologies and Adaptivity in Dialogue for Question Answering. AKA Press and IOS Press (January 2010)

35. Strachan, L., Anderson, J., Evans, M.: Pragmatic user modelling in a commercial software system. In: Proceedings of the 6th International Conference on User Modeling, Springer (1997) 189–200
36. Forbes-Riley, K., Litman, D.: A user modeling-based performance analysis of a wizarded uncertainty-adaptive dialogue system corpus. In: Proceedings of Interspeech. (2009)
37. Zukerman, I., Litman, D.J.: Natural language processing and user modeling: Synergies and limitations. User Model. User-Adapt. Interact. **11**(1-2) (2001) 129–158
38. Kaizer, S., Bunt, H.: Multidimensional dialogue management. In: Proceedings of the 7th SigDial Workshop on Discourse and Dialogue, Sydney, Australia (July 2006)
39. Brown, P., Levinson, S.C.: Politeness : Some Universals in Language Usage (Studies in Interactional Sociolinguistics). Cambridge University Press (February 1987)
40. Gruber, T.R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N., Poli, R., eds.: Formal Ontology in Conceptual Analysis and Knowledge Representation, Deventer, The Netherlands, Kluwer Academic Publishers (1993)
41. Oviatt, S., MacEachern, M., Levow, G.A.: Predicting hyperarticulate speech during human-computer error resolution. Speech Commun. **24**(2) (1998) 87–110
42. Paek, T., Horvitz, E., Ringger, E.: Continuous Listening for Unconstrained Spoken Dialog. In: Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP 2000). (2000)
43. Gebhard, P.: Emotionalisierung interaktiver Virtueller Charaktere - Ein mehrschichtiges Computermodell zur Erzeugung und Simulation von Gefuehlen in Echtzeit. PhD thesis, Saarland University (2007)
44. Ruckmick, C.A.: The Psychology of Feeling and Emotion. McGraw-Hill, New York (1936)
45. Morris, W.N.: Mood: The Frame of Mind. Springer, New York (1989)
46. Reithinger, N., Gebhard, P., Löckelt, M., Ndiaye, A., Pfleger, N., Klesen, M.: Virtualhuman: dialogic and affective interaction with virtual characters. In: ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces, New York, NY, USA, ACM (2006) 51–58
47. André, E., Dybkjær, L., Minker, W., Heisterkamp, P., eds.: Affective Dialogue Systems, Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004, Proceedings. In André, E., Dybkjær, L., Minker, W., Heisterkamp, P., eds.: ADS. Volume 3068 of Lecture Notes in Computer Science., Springer (2004)
48. Mehrabian, A.: Outline of a general emotion-based theory of temperament. In: Explorations in temperament: International perspectives on theory and measurement. Plenum, New York (1991) 75–86
49. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, MA (1988)
50. McCrae, R., John, O. In: An introduction to the five-factor model and its applications. Volume 60. (1992) 175–215
51. Minsky, M.: The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon & Schuster (November 2006)
52. Tosa, N., Nakatsu, R.: Life-like communication agent -emotion sensing character "mic" & feeling session character "muse"-. In: ICMCS '96: Proceedings of the 1996 International Conference on Multimedia Computing and Systems (ICMCS '96), Washington, DC, USA, IEEE Computer Society (1996)

53. Batliner, A.: Whence and whither: The automatic recognition of emotions in speech (invited keynote). In: PIT '08: Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems, Berlin, Heidelberg, Springer-Verlag (2008)
54. Rotaru, M., Litman, D.J., Forbes-Riley, K.: Interactions between speech recognition problems and user emotions. In: Proceedings of Interspeech 2005. (2005)
55. Litman, D.J., Moore, J.D., Dzikovska, M., Farrow, E.: Using natural language processing to analyze tutorial dialogue corpora across domains modalities. [67] 149–156
56. van Turnhout, K., Terken, J., Bakx, I., Eggen, B.: Identifying the intended addressee in mixed human-human and human-computer interaction from non-verbal features. In: ICMI '05: Proceedings of the 7th international conference on Multimodal interfaces, New York, NY, USA, ACM (2005) 175–182
57. Mehrabian, A.: Nonverbal Communication. Aldine-Atherton, Chicago, Illinois (1972)
58. Valstar, M., Pantic, M.: Fully automatic facial action unit detection and temporal analysis. In: CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, IEEE Computer Society (2006) 149
59. Tao, H., Huang, T.S.: Connected vibrations: A modal analysis approach for non-rigid motion tracking. In: CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, IEEE Computer Society (1998) 735
60. Keltner, D., Ekman, P.: Facial Expression of Emotion. In: Handbook of Emotions. The Guilford Press (2000) 236–249
61. Forbes-Riley, K., Litman, D.J.: Adapting to student uncertainty improves tutoring dialogues. [67] 33–40
62. Nelson, T.O., Narens, L.: Metamemory: A theoretical framework and new findings. In: G. H. Bower (Ed.) The Psychology of Learning and Motivation: Advances in Research and Theory. Volume 26. Academic Press (1990) 125–169
63. Sonntag, D.: On introspection, metacognitive control and augmented data mining live cycles. CoRR **abs/0807.4417** (2008)
64. Chu, S.W., ONeill, I., Hanna, P., McTear, M.: An approach to multi-strategy dialogue management. In: Proceedings of INTERSPEECH, Lisbon, Portugal (2005) 865–868
65. Studer, R., Ankolekar, A., Hitzler, P., Sure, Y.: A Semantic Future for AI. IEEE Intelligent Systems **21**(4) (2006) 8–9
66. Huang, T.S., Nijholt, A., Pantic, M., Pentland, A., eds.: Artifical Intelligence for Human Computing, ICMI 2006 and IJCAI 2007 International Workshops, Banff, Canada, November 3, 2006, Hyderabad, India, January 6, 2007, Revised Seleced and Invited Papers. In Huang, T.S., Nijholt, A., Pantic, M., Pentland, A., eds.: Artifical Intelligence for Human Computing. Volume 4451 of Lecture Notes in Computer Science., Springer (2007)
67. Dimitrova, V., Mizoguchi, R., du Boulay, B., Graesser, A.C., eds.: Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, Proceedings of the 14th International Conference on Artificial Intelligence in Education, AIED 2009, July 6-10, 2009, Brighton, UK. In Dimitrova, V., Mizoguchi, R., du Boulay, B., Graesser, A.C., eds.: AIED. Volume 200 of Frontiers in Artificial Intelligence and Applications., IOS Press (2009)