Chapter 2
Assessing the Quality and Usability of Multimodal Systems

To develop a user-friendly system, it is necessary considering quality aspects as early as possible in development process. Therefore, the evaluation of the subjective quality of such a system is required with a broad understanding of the concepts of quality and usability. Both constructs are related and share certain aspects but are not identical. Quality is defined by Jekosch [85] as the

> Result of judgment of the perceived composition of an entity with respect to its desired composition.

whereas the *desired composition* is defined as the "totality of features of individual expectations and/or relevant demands and/or social requirements". Hence, measuring the quality of a system means to measure the user's appraisal of the perceived characteristics of the system on the basis of the users' individual goal and background. This means, the user has to experience the interaction with the considered system in an interaction test to be able to compare this experience with personal desires and expectations. However, it is questionable, to which extend users consciously compare a system with a desired version and whether they even have an idea of a desired system design, as discussed by Wechsung [181].

Usability, on the other hand, is defined by ISO 9241-11 [80] as the

> Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

with *effectiveness* being "the accuracy and completeness with which specified users can achieve specified goals in specified environments", *efficiency* considering the "resources expended in relation to the accuracy and completeness of the goals achieved" and *satisfaction* being defined as "the comfort and acceptability of the system to its users and other people affected by its use". This is a very task-oriented definition with the focus on functional aspects as it was the most important topic in the early years of human-computer interaction research. In literature, other definitions of usability, which focus on other aspects, can be found (e.g. [49, 123]). Like in the definition of quality, a common criterion is the involvement of users. Thus, in order to understand what constitutes usability, users and their behavior must be

© Springer International Publishing Switzerland 2015

J. Seebode, *Emotional Feedback for Mobile Devices*, T-Labs Series in Telecommunication Services, DOI 10.1007/978-3-319-17193-7_2

understood. Hence, usability is not simply a system property, but refers to the interaction between a user and a system in the context of the respective tasks and usage situations. Accordingly, usability can be seen as one part of quality.

Satisfaction as one aspect of usability is defined in the ISO standard but a review by Hornbaek of 180 studies published in HCI journals and proceedings showed that satisfaction is the least frequently surveyed and most vague usability factor. Hence, it needs to be extended and standardized [77]. With increasing interest in *user experience* (UX), new concepts and methods have been developed, which focus more on the user satisfaction component. However, there is much discussion on how UX can be defined. A broad definition given by ISO 9241-210 [82] describes UX as

> A person's perceptions and responses that result from the use or anticipated use of a product, system or service.

According to Bevan [22], user experience can be seen in different ways. The first interpretation describes UX as

> An elaboration of the satisfaction component of usability [21].

In this thesis, this definition will be used as it is also the basis for the taxonomy described in Sect. 2.1.

The second interpretation differentiates UX from usability, as usability

> has a historical emphasis on user performance [145].

mostly neglecting the satisfaction component.

The third possibility is to understand UX as

> An umbrella term for all the user's perceptions and responses, whether measured subjectively or objectively [82].

which is the broad interpretation given by the ISO 9241-210 definition.

However, it is still not fully understood how performance measures and user perception correlate and how long-term usage influences satisfaction, even if the development of user experience over time has been in the focus of research (e.g. [94]). Hence, technical characteristics and user ratings should be carefully recorded and analyzed in the quality evaluation of multimodal systems. The following taxonomy gives an overview of quality aspects that describe multimodal interfaces.

2.1 Taxonomy of Quality Aspects of Multimodal Interfaces

A first step to understand multimodal interaction and to find suitable methods to investigate their quality was the development of a taxonomy of quality aspects of multimodal interfaces by Möller et al. [116]. This taxonomy is described in detail in its final version by Wechsung [181] and will be outlined in the following section. The complete taxonomy can be seen in Fig. 2.1. For this work, the taxonomy serves

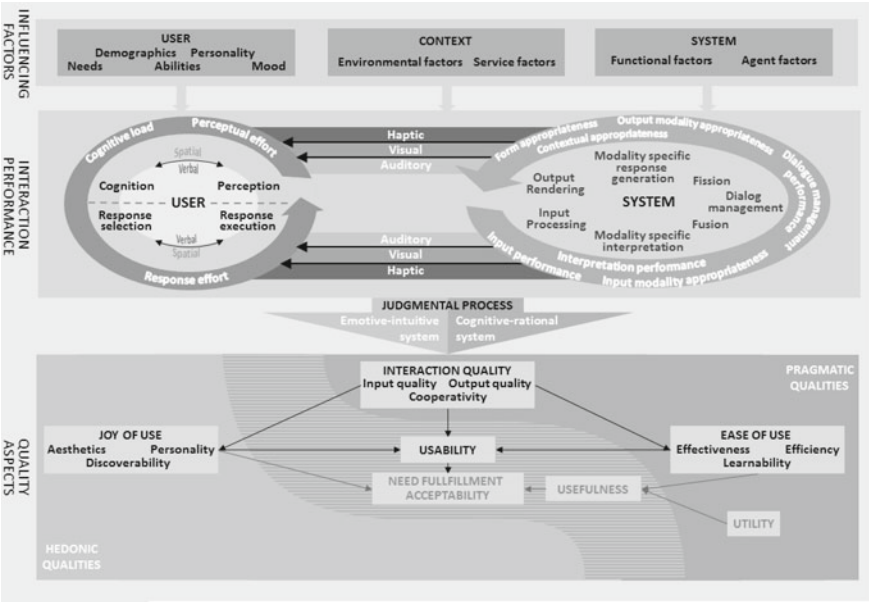

Fig. 2.1 Taxonomy of multimodal quality aspects, taken from Wechsung [181]

as starting point to localize and systematically address research gaps in the investigation of the quality of multimodal system feedback. The taxonomy consists of three layers showing the interplay between influencing factors, interaction performance and quality aspects.

The first layer contains the **influencing factors** including *user characteristics*, *context* and the *system* itself. User factors include all characteristics that describe the users and have an influence on their behavior and quality judgment like age, gender, cultural background and personality as well as motivational aspects, mood and needs. The context comprises the physical environment that can vary according to the situation in which the interaction takes place, and can be described, e.g. with background noise, lightning and the presence of other people or parallel tasks as well as service factors, referring to the availability of the system and related costs. The system itself is described in terms of functional capabilities (functional factors) and the characteristics of the system as an interaction partner (agent factors). All these factors influence the interaction behavior and thus, have an impact on the perceived quality [116].

The second layer of the taxonomy describes the **interaction performance** with the processes and behavior of *user* and *system*. On the user side, aspects of *cognitive load*, *perceptual effort* and *response effort* are important. The system performance can be described in terms of *input performance*, *interpretation performance*, *input modality appropriateness*, *dialogue management performance*, *output* *modality appropriateness, contextual appropriateness*, and *form appropriateness*. It has to be noted, that these parameters do not give direct information about the perceived quality, rated by users. They are therefore referred to as *indirect measures* [116] in contrast to *direct measures*, which are obtained by directly asking users. As a consequence, indirect data does not always correlate highly with users' perceptions. Thus, finding relations between direct and indirect data is the goal of some research (e.g. [120, 159]). Metrics to measure the system performance are given in Sect. 2.2.1.

The third layer comprises the **quality aspects**, which describe all important aspects of the users' experiences influencing quality ratings. As these ratings are based on the previous judgmental processes and depict the users' subjective experiences, it is indispensable to obtain this data directly by asking users. The judgmental process has been described with the involvement of two systems: a *rational system* and an *emotionally driven system* [54, 88]. Accordingly, a *pragmatic or ergonomic quality* and a *hedonic*[1] *quality* [67, 69] can be assigned to each interaction with a multimodal system. Thus, the quality aspects layer of the taxonomy is divided in hedonic and pragmatic quality aspects. Hedonic qualities refer to the joy-of-use component with its non-instrumental aspects like *aesthetics*, *personality* and *discoverability*, that educe fun and an interesting, engaging, or cool interaction.

The term *satisfaction* as stated in the definition of usability, is not named explicitly in the taxonomy, but is addressed by this joy-of-use component. Pragmatic qualities mainly refer to the ease-of-use component which comprises effectiveness, efficiency, and learnability and thus represents the classical concepts of usability. Additionally, *interaction quality* is part of the pragmatic qualities. It is positioned in the upper center because of the close relation to the interaction performance. Empirical results of Wechsung [181] indicated a strong relation between ease-of-use and interaction quality, thus these aspects were located within the pragmatic qualities and not between hedonic and pragmatic qualities. Interaction quality consists of *input quality*, *output quality* and *cooperativity*. The last pragmatic aspect is *utility*, which refers to the functionality of a system, i.e. the different functions that are provided by a system. In between hedonic and pragmatic qualities the concepts of *usability*, *usefulness* and *acceptability* (or need fulfillment) are located. *Usability* is determined by joy-of-use, ease-of-use, and interaction quality. It leads to *acceptability*, which refers to the actual usage of a system, and thus is a more economic measure as described by Möller et al. [116]. *Usefulness* contains usability and utility. Thus, a useful system offers the functions expected by a user, and helps to employ these functions easily. It has to be noted, that the concepts of usability, usefulness, and utility are often only fuzzily distinguished [95] and sometimes the terms are used synonymously.

[1] The term hedonic comes from the Greek word *hedone* which refers to the creation of pleasure.

2.1.1 *Emotions and Affect in the Context of Multimodal Interaction*

Emotions and affect are getting more and more important in the research of human-machine interaction as they are related to the more satisfaction-focused concepts like *joy-of-use* and *user experience (UX)*. Accordingly, one major topic of this thesis is the investigation of the affective impression of different types of system feedback.

There is a variety of emotion theories that weight the aspects of emotions differently. One way of defining emotions relies on the concept of basic emotions that are said to be universal. But there is neither a last answer about how many (varying between 2 and 18 [127]) nor which basic emotions there are. One popular representative of this direction is Ekman [132]. His basic thought is that emotions developed with evolution and are accompanied by distinct facial expressions.

Other theories think of an emotion as a construct composed in a space of a certain number of dimensions. But again, different theories postulate varying numbers of dimensions (mostly two or three) with diverse labels. Already Wundt, one of the founding fathers of psychology, proposed three dimensions to structure emotions [188]: *Lust (pleasure)*, *Spannung (tension)*, and *Beruhigung (inhibition)*.

Osgood et al. [128] pursued this basic concept. They asked participants to rate given words with pairs of adjectives on a semantic differential and found three dimensions which they named *evaluation*, *activity* and *potency*. These could also be reproduced for nonverbal stimuli. Following this work, Mehrabian and Russel [112] developed a similar but shorter semantic differential with 18 pairs of adjectives which could also be reduced to three dimensions, namely pleasure, arousal and dominance. Bradley and Lang aimed to develop a universal and easy method to directly measure the affective reaction of a person on a stimulus. They designed a pictogram-based questionnaire, the *Self-Assessment Manikin* [26], named SAM in the following. It will be described in detail in Sect. 4.1. The three SAM dimensions are linked to the basic emotional-motivational processes of approach and avoidance for pleasant versus unpleasant stimuli [97]. They determine the direction of a behavioral reaction or intention as well as its intensity (corresponding to the arousal scale). These basic dimensions have been identified in physiological reactions to stimuli, e.g. changes in heart rate, as well as in verbal descriptions of emotional episodes. It could also be shown that stimuli with emotional valence—positive or negative—attract automatic attention and are processed more intensely [25, 97, 98]. Thus, the emotional connotation of a system feedback message can help to increase its salience for the user while at the same time a mismatch between emotional valence and associated semantic meaning or function may cause irritation.
The terms *emotion*, *affect*, and *mood* are related and often used as synonyms, but there are differences to be named.

Scherer [151] defines an emotion as:

> An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism.

However, it is still discussed, what can be regarded as the most important factor defining an emotion: the bodily changes mentioned by Scherer, the attention shift towards an emotional stimulus, the intention to react or the conscious evaluation of the experience [155]. Also Otto et al. [129] state that there is much discussion on the definition of emotion and related concepts. As a working definition for this work following Schleicher et al., it is sufficient to say that an emotion is a multicomponent phenomenon, which consists of physiological, behavioral, attentional, and conscious components [155].

Important related concepts mentioned in this work are mood and affect. Affect is often used synonymous with emotion but the individual experience and its intensity is more important than for emotions. There is also a close relation to basic emotions mentioned [154]. Mood is more easily discriminable as it refers to more unspecific emotional states that may last longer but are less intensive than emotions [129].

2.1.2 Locating the Quality Aspects in Focus of this Work

The taxonomy presented in Sect. 2.1 offers an holistic overview of the relevant concepts related to multimodal interaction. As this thesis is not intended to give insights into all related aspects, this section provides an overview of the concepts that are further investigated and their localization in the taxonomy of multimodal quality aspects.

Figure 2.2 highlights the aspects that are in scope of this work. Aspects of the user (especially mood) and context (in terms of the environmental factors) are taken into account from the first layer as well as agent factors (in terms of feedback) which concern the system itself. As mentioned before, mood is related to emotions and affect and can influence the users' perception of a system. Context is a very broad term which comprises many different aspects. This thesis will give a first insight into effects of context on the perception of feedback.

The second layer of the taxonomy focuses on the interaction between user and system. One part is the user's perception of the system output. Additionally, the output generation on the system side is regarded with emphasis on contextual appropriateness. In this work, the perception of auditory, tactile, and auditory-tactile feedback is investigated. The different modalities and their possible manifestations as feedback messages are described in Chap. 3.

Between the second and third layer of the taxonomy, the judgmental process is located. As the affective impression of feedback messages is one big part of this work, the emotional-intuitive system is regarded mainly. Nevertheless, also functional aspects are taken into consideration, as the cognitive-rational system is important, too. Accordingly, the output quality is influenced by the affective impression assessed through the emotional system and by the functional connotation that is judged with the cognitive-rational system.

As mentioned before, the usability of a system is influenced, by hedonic and pragmatic qualities. Whereas the affective impression of a feedback messages refers

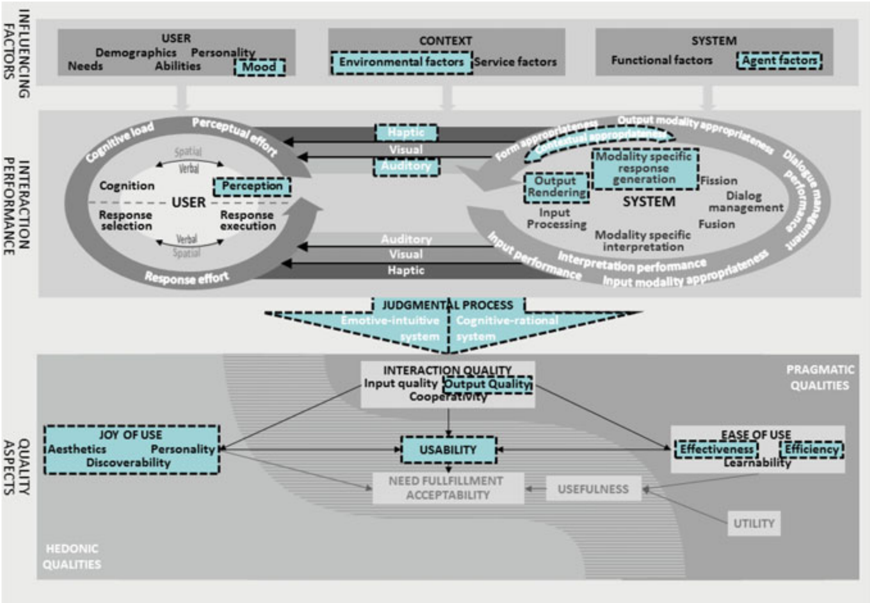

Fig. 2.2 Aspects in focus of this work are highlighted with *dotted lines* in the taxonomy of multimodal quality aspects (adapted from Wechsung [181])

to hedonic qualities and can influence the joy-of-use, functional aspects influence the ease-of-use components. The main focus of this thesis lies on the affective impression and functional connotation of feedback messages purely perceived, i.e. without an interaction paradigm (see Chaps. 5–7). Chapter 8 describes an interaction experiment, where also efficiency and effectiveness are assessed besides the affective impression and functional connotation.

2.2 Evaluation Methods

According to Sarodnick and Brau [147] an *evaluation* denotes "a systematic and preferably objective assessment, of a planned, ongoing or completed project". Dix et al. [49] name three main goals of an evaluation:

- to assess the extent and accessibility of the system's functionality
- to assess the users' experience of the interaction
- to identify any specific problems with the system

Additionally, they distinguish between a *formative evaluation* that is intended to improve a design and a *summative evaluation* that is used to verify whether the product is good enough.

For the evaluation of multimodal systems there have been few approaches validated, so far. Many commonly used methods and metrics are based on methods developed for spoken dialogue systems (SDS). An example of such an approach for the evaluation of multimodal systems is the PROMISE[2] framework [15]. It expands the PARADISE[3] framework [180] which was developed for the evaluation of SDS. The approach of PROMISE is to weight individual recognition modules differently[4] and to give appropriate metrics for the various inputs and outputs. Nevertheless, there are only very few applications and studies applying PROMISE (e.g. [14]). Recently published work by Kühnel [90] describes a set of interaction-parameters for multimodal interaction. In combination with the recent work by Wechsung [181] on the empirical evaluation of multimodal interaction, a holistic view can be obtained.

In the following sections, an overview of established evaluation methods, which can be used for multimodal systems, is given. The methods used in this thesis are described further. First, interaction parameters and their suitability for the evaluation of multimodal systems are discussed, followed by an overview of empirical methods involving participants (either experts or potential users). The chapter ends with an overview of methods to measure affect and emotional states of the users.

2.2.1 Interaction Parameters

Measurable interaction parameters are used to investigate aspects of the *interaction performance* of a system, which have been described in Sect. 2.1. These parameters can be assessed either for the whole system or considering the individual system components. The quality of the overall system depends on the performance of the individual components as well as their interplay [50]. Especially for multimodal systems, so far few metrics have been investigated in detail and much research is still needed to examine the relation of different modalities. Nevertheless, recent work by Kühnel [90] has shown that metrics developed for speech dialog systems can be extended for multimodal systems. Based on the PARADISE approach, she defined a set of interaction parameters that was able to afford reasonable prediction rates (i.e. with >50 % accuracy) for user judgements. These parameters can be characterized in the following categories:

Dialogue- and communication-related parameters: These parameters deal with the whole dialogue between system and user. They commonly contain time-related metrics like *dialogue duration* or *turn duration* [63] and parameters like the *number of interaction steps* or the *number of*

[2] PROcedure for Multimodal Interactive System Evaluation Framework.

[3] Paradigm for Dialogue System Evaluation Framework.

[4] For example gesture recognition often is more accurate than speech recognition, since the “vocabulary” is smaller. However, an improvement of the speech recognition module is more likely to lead to a higher quality rating of the whole system and should therefore be weighed higher [15].

words per turn. For systems with parallel multimodal input, also the *modality choice* can be assessed [121].

Task-related parameters: These parameters describe for example *task success*. Hence, they are meaningful only for task-oriented systems. They can be assessed for a whole dialogue or for smaller sub-tasks or interaction steps, leading to a concept like *request success* [152]. In interaction studies, the tasks have to be pre-defined with the state of a successful task completion.

Parameters for input: These parameters are well-known for SDS, quantifying e.g. *recognition and error rates* of the speech recognition module. They can be easily adopted for other input modules, keeping in mind, that other modalities like touch screens are less error prone. Nevertheless, it is not always clear which kind of error has more influence on the subjective system quality. Additionally, *multimodal error rates* can be calculated for the complete multimodal system [90].

Parameters for output: For system output, aspects like *delay* and *synchrony* are important. For systems with corresponding output modalities the *lag of time* between those can be measured as proposed by Beringer et al. [15]. Furthermore, e.g. for SDS the TTS quality can be measured as proposed by Möller et al. [117]. For auditory and tactile system output, parameters describing signal quality such as *noise*, *distortions*, and *frequency response* can be assessed.

Meta-communication-related parameters: Errors can occur in every interaction. Thus system developers should carefully think of methods to solve interaction problems and to give users the possibility to ask for help. Metrics to be assessed in this category are e.g. the *number of system error messages* or the *number of help requests* [179].

Cooperativity-related parameters: The term *cooperativity* has first been used for SDS by Dybkjaer et al. [51]. The parameter *contextual appropriateness* has been proposed by Simpson and Fraser [165]. It is measured by judging each system utterance as to whether it violates one or more of Grice's conversation maxims [65].

This description is neither exhaustive nor detailed because that would go beyond the scope of this thesis. More examples and details can be found in [90] and the other literature referred to above. Depending on the goal of the evaluation and the system itself, a set of meaningful parameters has to be selected for measurement. In the Interaction Study described in Chap. 8 mostly durations and task-related parameters were measured.

2.2.2 Evaluation Methods Involving Participants

Since the sole measurement of interaction parameters is not sufficient to obtain a complete knowledge about the usability of a system, it is necessary to conduct an evaluation with participants. Subjective quality and usability evaluation methods can be distinguished between methods that are conducted with experts (also referred to as *analytical evaluation methods*) and methods which use naive participants to test a system (also called *empirical evaluation methods*) [76]. Both methods are often applied in combination or sequentially. With the sequential approach it is possible to detect the biggest usability problems in early development phases with only a few experts, and performing more time-consuming and complex tests with potential users with sophisticated mock-ups or working prototypes later in the development process.

2.2.2.1 Expert Evaluation

The efficiency of an expert evaluation highly depends on the experience of the consulted experts. It can be distinguished between *usability experts*,[5] *domain experts*[6] and *usability and domain experts*. It has been shown that *usability and domain experts* in comparison to domain or usability experts can identify most problems [123].

In the following section two commonly used methods of expert evaluation are described first: the *Heuristic Evaluation (HE)*, the *Cognitive Walkthrough (CW)*, followed by a short description of a less common group of methods called *Model-based Evaluation (ME)*.

Heuristic Evaluation

A frequently used method is the Heuristic Evaluation (HE)[7] which is a method of *discount usability engineering*. This term was introduced by Nielsen as a pragmatic and cost-effective approach due to the argumentation that extensive empirical usability evaluation is too expensive, too difficult and would take too long [124].

During an HE, experts examine the system individually or in a group, taking into account a list of so-called heuristics. The aim is to find as many potential usability problems as possible. In the original version [124] it is required, that each evaluator examines the interface on its own by interacting with the interface several times. Only after completion of all reviews, the experts communicate with each other to find an

[5] These experts are meant to have a lot of experience in the field of usability.

[6] These experts have much experience in the area of to the system under investigation (for example, experience with speech recognition and synthesis for the evaluation of a spoken dialog system).

[7] The term *heuristic* is derived from the Greek "heureskein" which can be translated as "to find" or "to explore" [76].

aggregated judgment. The objective is to get independent and undistorted judgments of the individual experts. The result is a list of usability problems that can be classified by assigning them to individual points of the heuristic used. In addition, the problems found should be justified and further described in order to provide solutions and design recommendations. A prioritization of the different problems with respect to their frequency or influence on the interaction is also possible. According to Nielsen, approximately three to five experts are needed to uncover about 60–70 % of the problems in general and more than ten experts would hardly improve the results [124], which is still being discussed in the research community [181]. Although the approach was introduced by Nielsen, by now there exists a variety of different heuristics. Usually they rely on common usability principles and relate to a specific type of system. There have been heuristics specifically designed for spoken dialog systems [19], but to the authors knowlegde not for multimodal systems.

The HE has the great advantage to be fast, easy to apply and cheap. Furthermore, the HE is applicable throughout the entire design process, even with a not yet fully implemented system. The experts can evaluate the present system design based on a description or mock-up. However, it is arguable how well the experts are able to put themselves in the role of real users, even if this is a problem of most expert evaluation methods.

Cognitive Walkthrough

The Cognitive Walkthrough (CW) is a task-oriented expert evaluation method that is based on theories of learning by discovery and research on problem solving [186]. As with the HE, the aim of a CW is to find potential usability problems by experts. The experts—in most cases designers, usability professionals or psychologists—analyze the planned interaction with the system stepwise based on given tasks. For this purpose, it has to be defined:

- What is the usage context?
- Who are the target users?
- What is the ideal interaction path to perform each task?

With this approach, the experts generate a list of usability problems for the given tasks and user groups. Advantageous of the CW is the independence from real users and a fully functioning system. The CW may also help designers to see the system from the perspective of a potential user and thus to redefine users' goals and assumptions about the system [76]. The disadvantage here is that the outcome is just a list of the problems without possible solutions, depending on how extensively the experts comment on the problems found. Furthermore, a CW may provide biased results depending on the task and expert selection and with a complex system the procedure can become very tedious [76].

Model-Based Evaluation

With Model-based Evaluation methods (ME) the aim is to predict user behavior during the interaction with a system on the basis of user models [172]. In its simplest types, only duration and accuracy of arm or mouse movements (*Fitt's Law*) or execution times using the keyboard (*Keystroke Level Model* or *Goal Operators Methods Selection Rules (GOMS)*) are predicted. More complex cognitive processes can be modeled with the help of theories in the field of cognitive psychology [49] or expert knowledge from the respective fields of application. Examples of these cognitive modeling approaches are task analysis like the *Hierarchical Task Analysis* [9], the *Adaptive Control of Thought-Rational (ACT-R)* approach [8], and the *State and operator resultant model (SOAR)* [122]. Particularly, the latter methods are mainly used in academia rather than in the applied field.

As the other expert methods, also the ME can be applied very early in the design process. Additionally, it provides a deep insight into the behavior of users. The most important disadvantage of the ME is the time required, as each task must be analyzed step-by-step. In addition, a very high degree of expertise is necessary to sustain valid results [76]. Accordingly, it demands much effort even to model simple systems or small applications. Nevertheless, ME can be very useful for usability reviews of applications with a small user group or if an accurate assessment of all possible consequences of every action is required (e.g. in a plane, power plants).

2.2.2.2 Empirical Evaluation

The empirical evaluation is usually carried out as a test with participants who act like real users. With those tests it is possible to obtain data about the system itself (see Sect. 2.2.1 on *interaction parameters*), the user behavior and the perceived system quality.

One advantage of empirical evaluation methods is, that the problems found are really occurring and not only problems, which were predicted by experts. Additionally, performance data assessed using log files can be correlated with the received user judgments. Thus, the impact of problems on user perception is determinable. The main disadvantages of empirical evaluation methods are the high costs of time and money.

Depending on the aim of the evaluation, the practical test can be conducted differently. In a research environment, certain hypotheses are first derived from literature review, and are then checked using a laboratory setup under controlled conditions. A practical approach, however, has the main goal to uncover as many problems as possible.

For both approaches, it is important to design and describe certain tasks for the test participants.[8] The thorough selection of tasks must be done prior to the test. Likewise,

[8] There is also the possibility to let users freely explore a system, but especially in the research context, this is less common.

the order and presentation of the tasks needs to be carefully thought through. In the test itself, the given tasks are accomplished by the participants under observation by the experimenter. After completing each single task or a list of tasks, the participants are asked to rate the system and the interaction to assess their perception. This can be done by means of questionnaires or interview techniques. Many questionnaires have been developed for specific systems and are hardly transferable, especially for multimodal systems. In addition, their reliability[9] and validity[10] must be checked. For this reason, much effort has been spent solely on the development of appropriate questionnaires for multimodal dialogue systems [181].

One very general method designed to determine user judgments concerning interactive products is given by the AttrakDiff [67]. It presents an attempt to investigate quality features beyond classical usability (i.e. perceived pragmatic quality). Thus, also aspects of hedonic quality and overall attractiveness are examined. The original AttrakDiff contains 28 entries, which are presented in the form of a semantic differential. Users evaluate the system using opposite pairs of adjectives (e.g. “simple—complicated”) on a seven-point scale. The 28 items are assigned to four factors or dimensions:

- **Pragmatic Quality (PQ)** describes the usability of the system and the degree to which users can achieve their goals with the system.
- **Hedonic Quality-Stimulation (HQ-S)** illustrates how the system can support users in further developing themselves by offering new and exciting concepts.
- **Hedonic Quality-Identity (HQ-I)** indicates how well users can identify themselves with the system.
- **Attractiveness (ATT)** is a global dimension, which describes the overall perceived quality.

The dimensions of pragmatic and hedonic quality are independent and contribute equally to the overall attractiveness judgment. Furthermore, in contrast to the hedonic quality, the pragmatic quality highly correlates with the perceived effort when interacting with a system.

A new and shorter version, the AttrakDiff Mini [68], has also been developed. It contains 10 items: four of them measuring pragmatic and four measuring hedonic aspects (collapsing the two hedonic dimensions of the original AttrakDiff), and single additional items to rate the goodness (i.e., “bad—good”) and beauty (i.e., “ugly—beautiful”) of a system or product directly. The AttrakDiff Mini is used in the Interaction Study, presented in Chap. 8.

A recent approach to design a questionnaire especially for multimodal systems used the AttrakDiff as a basis: the *MultiModal Quality Questionnaire (MMQQ)* developed by Wechsung [181]. Its theoretical foundation is the taxonomy described

[9] Reliability describes the accuracy and the extend to which an instrument (e.g. a scale or a questionnaire) yields the same results in a repeated test or in sub-parts of a test.

[10] Validity describes, whether a method or measurement (e.g. a questionnaire) does actually measure what it is supposed to do.

in Sect. 2.1 and aim of its development was to cover all quality aspects of multimodal interaction. Another recent approach to measure user experience is the modular meCUE questionnaire developed by Minge et al. [114].

2.2.3 Measuring Affect and Emotions

A very popular rating scheme for assessing the affective impression of different stimuli is the aforementioned *Self-Assessment Manikin (SAM)* [26]. It consists of three pictorial scales representing the three affective dimensions: valence, arousal and dominance. As the SAM is one of the mayor questionnaires used in this work, a more detailed description is given in Sect. 4.1.

Another tool to measure specific predefined emotions is *PrEmo* [44] that uses fourteen expressive cartoon animations to measure the emotions evoked by a product. A disadvantage of PrEmo is, that the method is only validated for non-interactive products, so far [44].

An easy way to assess positive aspects during computer usage is given by the "Joy of Use Button" presented by Schleicher et al. [156]. The rationale behind it is to provide a simple tool for users to measure *joy-of-use* immediately in the moment when it occurs during the interaction.

There have also been approaches to correlate affective reactions with psycho-physiological measures like electrocardiography (ECG), electromyography (EMG), electroencephalography (EEG) and electrodermal activity (EDA) [154]. As facial muscle activity is not completely controllable, EMG measures are widely used to measure the valence of a situation (positive or negative) [155]. Bradley and Lang [27] observed high correlations between facial muscle activity and valence ratings for the corrugator muscle (responsible for frowning) and the zygomatic muscle (activated for smiling). Furthermore, EMG measures have been used to asses the usability of a system, which is a less straight forward approach with rather unspecific results based upon EMG alone [108]. Moreover, physiological measures are resource-intensive as it takes much time and expertise to build up an experimental setup to measure and also synchronize even peripheral physiological measures like ECG and EMG [154].

Springer
http://www.springer.com/978-3-319-17192-0

Emotional Feedback for Mobile Devices
Seebode, J.
2015, XIV, 131 p. 29 illus., 13 illus. in color., Hardcover
ISBN: 978-3-319-17192-0