Steps in Identifying Interaction Design Patterns for Multimodal Systems

Andreas Ratzka

IMIK, Universität Regensburg
93040 Regensburg
Andreas.Ratzka@sprachlit.uni-regensburg.de

Abstract. The context of this work is usability engineering for multimodal interaction. In contrast to other work that concentrates on prototyping toolkits or abstract guidelines, this research focuses on user interface patterns for multimodal interaction. Designing multimodal applications requires several skills ranging from design and implementation. Thus, different kinds of patterns (from architecture patterns to user interface patterns) can be applied to this field. This work focuses on user-task near user interface patterns. At first, a traditional approach of modality selection based on task- and context-based rules is presented. Next, a twofold process of pattern mining is presented. In the first phase, pattern candidates are derived top-down from proven knowledge about how multimodality enhances usability. In the second phase, literature is mined for real solutions to underpin these pattern candidates and find new ones. Along with this, relationships between patterns are depicted.

1 Introduction

The context of this work is usability engineering for multimodal interaction. Traditional approaches in this field focus on prototyping [15, 16, 30] or decision support for requirements analysis and work reengineering [6, 9, 32]. The later stages in the usability engineering lifecycle, i.e. design standards and detailed design, are only marginally covered by those decision support systems.

The idea of this work is to apply the concept of design patterns to the field of multi-modal interaction. A design pattern is a rule connecting a common design problem with a proven solution and a description of the contexts and conditions in which this pattern is applicable [8, 17].

The idea of patterns originates from architecture [1, 2] but has gained popularity mainly in different fields of computing such as object orient programming [18], software architecture [10] and user interface design [7, 40, 41, 42].

A good pattern provides a solution which cannot be derived from general guidelines using trivial mapping rules. A pattern is a context-specific design rule that discusses why other apparent solutions are not applicable in this context. This is done in pattern sections titled *forces* – to discuss the goal conflicts impeding simple and obvious solutions – and *consequences* – to discuss how the goal conflicts are resolved by the proposed solution and which new problems might arise.

Multimodal interaction has not yet reached wide-spread market penetration. Nevertheless, after almost thirty years of research, several demonstration systems have been designed. Recurring problems have lead to solutions which were reused successfully in subsequent projects so that these solutions can be identified as interaction design patterns [34].

Designing multimodal systems requires a lot of skills comprising among others software architecture, implementation techniques, speech and screen design, and task modelling. Each of these fields can be supported by different kinds of patterns such as the (implementation-near) architecture patterns PAC, MVC and Blackboard [10] or (user-task-oriented) user interface patterns such as those described in [40, 41].

This work focuses on patterns of the latter (user-task-near) category. Even within this group, one can distinguish different levels of granularity. This paper describes on the one hand higher level patterns that are based on the general principles of the multimodal design space (patterns of multimodal combination and multimodal adaptation), as well as more concrete use case specific patterns on the other hand [36, 37].

Similar approaches for multimodal interaction are rare. Only the work described in [19] goes in the same direction and identifies patterns for multimodal interaction. However, that work emphasises formalisation and avoids direct links to already existing "traditional" user interface patterns. This work, by contrast, identifies specific multimodal interface patterns and attempts to put them in relation to traditional, more general user interface patterns.

This paper illustrates first a simplified approach of modality selection which is based on design rules that are derived from modality theory and interaction constraints. The designer selects appropriate modalities according to the requirements of the target application. This approach results in propositions such as "use modality A", which are helpful during the first phases of usability engineering. But it lacks more detailed speech and screen design recommendations. This work assumes that patterns can complement this gap and provide decision support across all design phases.

The following sections describe the process of mining user interface patterns which consists of two temporally overlapping phases.

In the first (top down) phase, user interface patterns are derived from general properties of the multimodal interaction design space. In the second (bottom up) phase concrete use cases are discussed. This paper focuses on mobile applications, discusses, how traditional user interface patterns [40, 41, 42, 43] can be applied, and identifies new user interface patterns that build specifically on multimodal interaction techniques.

Patters are not standing alone but are mutually interrelated and form a pattern language [25]. Relationships cover typically usage (pattern A makes use of pattern B) and refinement (pattern A is refined by pattern B). Beyond relationships among specifically multimodal user interface patterns, this paper illustrates relationships between multimodal and traditional user interface patterns such as those found in [40, 41, 42, 43].

2 Traditional Approach of Design Support: Modality Selection based on Task Properties and Context of Use

Traditional approaches such as modality theory and modality properties [6], interaction constraint models [9, 32] and other guidelines for multimodal interaction provide solutions for design problems. This section exemplifies modality selection according to task properties and context-based constraints.

2.1 Modality Selection according to Task Properties

The first step in designing multimodal interactive systems is to elicit interaction modalities that are appropriate for the current task. One starting point are the modality properties described in [6], which tackle following issues:

- **Required interaction channels** (Spoken language is conveyed auditively, written text visually)
- **Salience** (Auditive signals are more attention catching than visual ones)
- **Local selectivity** (visual data are perceived only if they are paid attention to)
- **Degree of user control** (static modalities like written text allow more user control over pacing than dynamic modalities such as videos or spoken text)
- **Learning requirements** (arbitrary modalities such as newly defined symbols require more learning efforts than those building upon existing conventions)
- **Expressiveness** (analogous modalities such as graphics are preferred for conveying spatial relationships whereas linguistic modalities like text convey conceptual information such as detailed descriptions better).

Rules taken from modality theory and modality properties [6] are universally valid and expected to be stable even for novel interaction techniques. Nevertheless a concretisation for each individual project, for currently available modality combination is needed. Figure 1 shows an exemplary task-modality matrix, which gives the user advice on modality selection.

Input							Output					
	Speech Input	Typing	Handwriting	Pointing	Free Gestures	Eye movements		Auditive Output	Visual Output	Haptic Output	Motor Input	Speech Input
Sketching	–	–	–	+	?	?	Urgent Information	+	?	?		
Graphic manipulation	?	?	?	+	?	?	Highly current information	+	?	?		
Selecting (small sets)	+	?	?	+	+	?	Status information	–	+	?		
Selecting (large sets)	+	+	?	?	?	?	Private information	–	+	+	+	–
Text input	?	+	+	?	?	?	Security relevant visual primary task	+	–		–	+

Fig. 1. Exemplary Modality Selection Criteria based on Task Characteristics.

2.2 Interaction Constraints based on Context of Use

After selecting (several alternative) task appropriate modalities, the designer has to check further interaction constraints imposed by user characteristics, device characteristics and the environment [9, 32]. These additional constraints can be cast into similar problem-solution matrices such as the one for task characteristics. However, it is difficult for the designer to keep track of a bunch of several constraint matrices all at once.

Instead, these additional interaction constraints are presented in an (exemplary) contradiction matrix. The columns of this matrix contain cases that encourage the use of an individual interaction modality whereas the rows are listing those cases that discourage the respective modality. The designer first checks which interaction modalities are most appropriate for the tasks to be supported by the system. Then he checks whether for each individual candidate modality the factors listed in the columns outweigh the factors listed in the rows and contrasts these results for each interaction modality. Roughly speaking, the fields near the matrix diagonal (crossed out in our examples) mark cases of conflicting usability goals.

Output constraints

			Speech output preferred						Audio tones preferred					Visual text preferred			Graphics preferred		
			Illiterate people	Visually impaired people	Car user interfaces	Phone user interfaces	Smartphones, PDAs	Bad lighting conditions	Visually impaired people	Car user interfaces	Phone user interfaces	Smartphones, PDAs	Bad lighting conditions	Hearing impaired people	Noisy environments	Public, crowded environments	Hearing impaired people	Noisy environments	Public, crowded environments
Speech output	discouraged	Hearing impaired people	X	X	X	X	X	X											
		Noisy environments	X	X	X	X	X	X											
		Public / crowded environments	X	X	X	X	X	X											
audio tones	discouraged	Hearing impaired people							X	X	X	X	X						
		Noisy environments							X	X	X	X	X						
		Public / crowded environments							X	X	X	X	X						
Visual text	impossible	Illiterate People												X	X	X			
		Car user interfaces												X	X	X			
	discouraged	Visually impaired people												X	X	X			
		Bad lighting conditions												X	X	X			
Graphics	discouraged	Visually impaired people															X	X	X
		Car user interfaces															X	X	X
		Bad lighting conditions															X	X	X

Fig. 2. Exemplary Output Constraints

Input Constraints

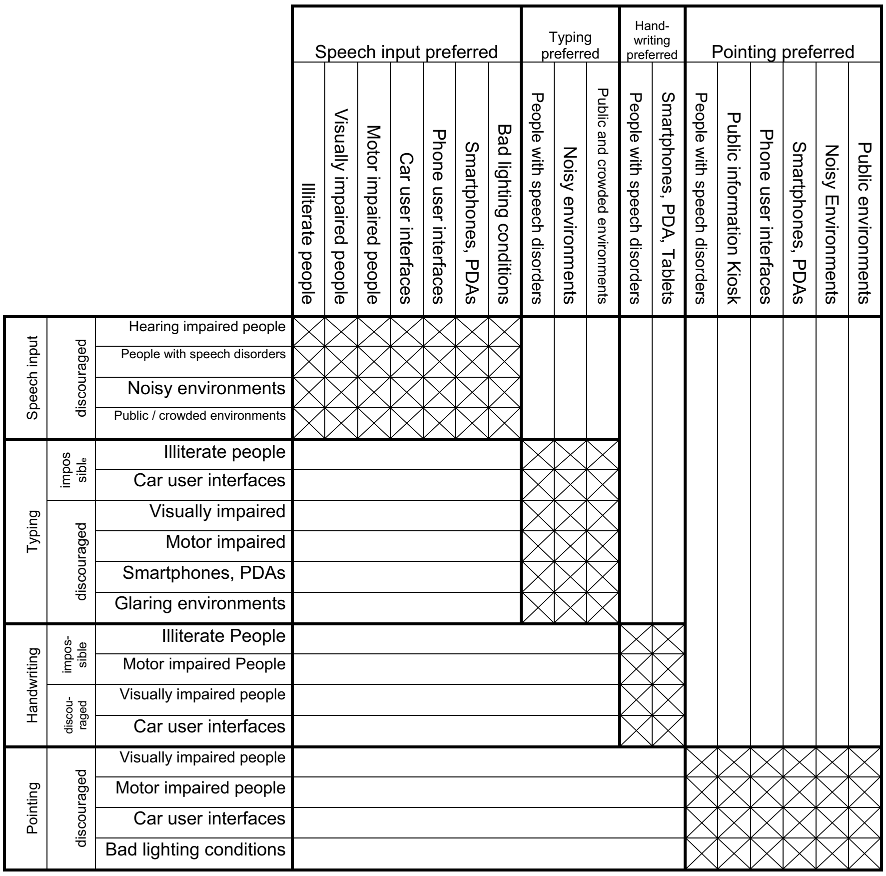

Fig. 3. Exemplary Input Constraints

2.3 Shortcomings of this Approach

This traditional approach is valuable for the first steps of user interface design. Nevertheless it lacks detailed design recommendations on how several modalities have to be combined and coordinated, which requires more detailed guidelines.

This work assumes that patterns are a valid approach to provide design support across all phases of user interface design. The next sections outline the twofold process of identifying user interface patterns for multimodal interaction. This process is both top-down – based on general principles of multimodal interaction – and bottom-up – based on real world examples of multimodal interactive systems.

3 Deriving Patterns from Generic Principles of Multimodal Interaction

According to [31] multimodal interaction can be classified along several orthogonal dimensions. The main dimensions of fusion (content related vs. unrelated) and parallelism (temporally overlapping vs. sequential) lead to four major classes of exclusive, alternating, concurrent and synergistic multimodality.

The potential of multi-modal interaction lies in enhanced flexibility, naturalness, robustness and interaction performance. This can be achieved via suitable modality combinations as well as via selection of appropriate interaction modalities, that is via adaptation during runtime.

The CARE properties [14] define classes of modality combination in multimodal interactive systems:

- **Equivalence**: One piece of information can be exchanged via several modalities alternatively
- **Specialization:** One piece of information can only be exchanged via one interaction modality
- **Redundancy:** One piece of information is conveyed via several interaction modalities in a redundant way.
- **Complementarity:** Several connected pieces of information are conveyed via several mutually complementing modalities

3.1 Patterns for Modality Combination

Modalities are combined to minimise task interference, maximise information throughput, disambiguate distorted input (and output) signals, optimise saliency and assure usability across diverse and varying contexts of use.

Patterns identified in the context of modality combination are:

- Audio-visual Workspace (makes use of complementarity)
- Audio-visual Presentation (makes use of complementarity)
- Redundant Input (makes use of redundancy)
- Redundant Output (makes use of redundancy)

Following section outlines the pattern *Redundant Input* in some more detail.

Redundant Input

Context

Communication channels might be unpredictably distorted due to bad lighting conditions, background noise, technical (network) problems or disabilities such as speech, motor or perception disorders.

Problem

How to assure input when communication channels are distorted in an unforeseeable way?

Forces

- The system can be configured to use interaction modalities that are less affected by channel disorders but in some cases all available interaction channels are distorted to some degree. Consider following scenarios:
  - How to support hands free tasks in noisy environments?
  - How to interact with motor-impaired users in loud environments?
  - How to interact with people with speech disorders in a hands-free scenario?

Solution

Combine several interaction channels in order to make use of redundancy. Input coming from several channels (visual: e.g. lip movements, auditive: e.g. speech signal) should be interpreted in combination in order to reduce liability to errors.

Consequences

- Even if several channels are distorted the distortion rarely affects exactly the same pieces of information. Combining sound pieces of information from several channels some distorted parts can be reconstructed:
  - In loud environments, speech recognition performance increases significantly when audio-signals are combined with visual signals (from lip movements).
  - Multimodal speech recognition can increase recognition performance for accent, exhausted and disordered speakers.

Rationale

Independent disturbances of different channels rarely affect the same aspects of the content. That's why for instance audio-visual speech recognition which combines acoustic signals and lip movement analysis leads to better recognition performance than unimodal speech recognition [5, p. 24 f.]:

Plosives ([p], [t], [k], [b], [d], [g]) sound similar and are likely to be confused when sound quality is low. At the same time these phones have distinctive lip shapes such as open lips (in the case of [g] and [k]) vs. initially closed lips (in the case of [b] and [p]). Lip shapes may differ for some similar sounding vowels, too.

Distortions rarely affect both the recognition of (acoustic) phonemes and corresponding "visemes" in the same way. Fusion algorithms allow to combine sound pieces of information from several channels to reconstruct distorted parts.

Known Uses

This variant is manifested in very different application areas including among others data input (audio-visual speech recognition), person identification [39], emotion recognition [44].

3.2 Patterns for Modality Adaptation

Systems that are used by different users subsequently (changing users), by individual users extensively (growing user expertise), in different or changing environments, or with changing degrees of service availability (changing network bandwidth) have to be adapted to these unforeseeable context factors. Adaptation can be done automatically (channel analysis, user modelling, etc.) or initiated by the user (changed behaviour or explicit configuration). Based on these aspects, following patterns, which require the presence of *equivalent* modalities, were identified (for a detailed description cf. [37]):

- Multiple Ways of Input
- Global Channel Configuration
- Context Adaptation

4 Identification of Multimodal User Interface Patterns based on Real World Examples – Illustrated by Mobile Systems

Patterns are never inventions by their authors but always relate to – at least three – successful examples of system design [8]. Among several use-cases such as mobile interaction, interactive maps, graphic design applications and systems for augmented dual-task environments, mobile systems are selected for detailed discussion, underpinning of pattern candidates and pattern identification.

Examples for multimodal mobile interaction are personal assistants for e-mail and web access such as *MiPad* [22], *Personal Speech Assistant* [13], tourist guides and city information systems such as *SmartKom mobile* [26], *MATCH* [21, 24], *MUST* [3] or *COMPASS* [4].

4.1 Pattern Discussion based on Use-case Aspects

Multimodal mobile systems and smartphones make use of spoken commands to avoid the necessity of deep menu navigation for starting programs, placing phone calls etc. This new user interface pattern is called *Voice-based Interaction Shortcut* [36] and can be used in diverse interaction scenarios.

Starting an Application

The pattern *Hub and Spoke* [41] is an appropriate approach for organising applications on mobile devices. Each one of the most important applications is easily reachable from the main page. At the same time, when leaving an application, you return to the main page as well. This way, orientation can be granted despite the lack of space.

Additionally, mobile devices usually provide so called *quick launch buttons* to start the four or five most common applications with one press. This can be seen as an extension of *Hub and Spoke*.

The above mentioned pattern *Voice-based Interaction Shortcut* can be applied for launching applications in one interaction step. This way, the desired program can be started without the need for the current display to include a direct link to this application.

List Selection
List selection is another application area for the pattern *Voice-based Interaction Shortcut*. Instead of scrolling through lists or poking on a screen keypad the user can simply speak the desired list item.

Structured Text Input
Text input can be facilitated using the pattern *Autocompletion* [41]. The user only has to input some letters until the list proposed by the system includes the desired entry. Similarly list selection in very large lists can be alleviated by applying the pattern *Continuous Filter* [42] allowing the user to enter the first letters of an entry until no scrolling is necessary any more.

In some cases structured input is necessary. Think of web *form*s [40] or e-mail messages. The user has to select an input field and then enter textual information. Some input fields can be enriched with a *Dropdown Chooser* [41] to offer list selection instead of text input. If this *Dropdown Chooser* is enriched with the pattern *Voice-based Interaction Shortcut* in the context of structured input forms we receive as a result the new multimodal user interface pattern *Speech-enabled Form* [36].

The mobile multimodal organiser MiPad makes use of this pattern as it allows the user among others to create e-mails via combining pen input and spoken language the following way: When the user selects the receiver field, a recognition vocabulary consisting of contact items is selected and speech recognition is activated. When the user selects the subject or message field, a free-text recognition vocabulary is selected instead.

The user's tapping with the pen onto the input field is used to activate the speech recogniser. This is important because speech recognition must not be active all time, otherwise background noise, private speech, respiration and harrumphing could lead to undesired results. Instead, activating the recogniser via tapping and deactivating it after input or a certain period of time can avoid this problem. Thus, *Speech-enabled Form* makes use of Tidwell's [41] pattern *One-off Mode*.

Implementation techniques supported by XHTML+VoiceXML [23] enforce this *Speech-enabled Form* paradigma.

Avoiding Recognition Errors
Mobile messaging systems [27] and car navigation systems [29] deal with large vocabularies that can lead to poor speech recognition performance. To improve dialogue quality some systems offer the user not only to re-speak the misrecognised word or phrase but to select it from a list – via pointing, speaking the line number or re-speaking with additional attributes. This change of input technique is important as it avoids endless error-correction loops. The presentation of the n-best list in a *Dropdown Chooser* [41] which allows the user to correct initially spoken words via pointing is a new multimodal user interface pattern called *Multi-modal N-best Selection* [37].

Other systems propose the user to spell or type the first character(s) of the item/name to be input. This way the size of speech recognition vocabulary can be reduced which results in more robust recognition performance. This combination of *Continuous Filter* and *Voice-based Interaction Shortcut* results in the new pattern *Spelling-based Hypothesis Reduction* [37].

Both *Multi-modal N-best Selection* and *Spelling-based Hypothesis Reduction* are specialisations of the above mentioned pattern *Redundant Input*.

4.2 Summary of Identified Patterns

Following patterns were identified for mobile multimodal interaction:

- Voice-based Interaction Shortcut
- Speech-enabled Form
- Multimodal N-best Selection
- Spelling-based Hypothesis Reduction

Pattern Relationships

The four main patterns identified in this paper are in close relationship to one another: The pattern *Voice-based Interaction Shortcut* is used by *Multi-modal N-best Selection* as well as by *Speech-enabled Form*. Speech-enabled Form, a refinement of Tidwell's [40] *Form*, makes use of *Spelling-based Hypothesis Reduction* and *Multi-modal N-best Selection* as well as of Tidwell's [41] *One-off Mode*. *Multi-modal N-best Selection* makes use of Tidwell's [41] *Drop-down Chooser*. *Spelling-based Hypothesis Reduction* uses the pattern *Continuous Filter* [42]. Following figure illustrates these relationships visually.

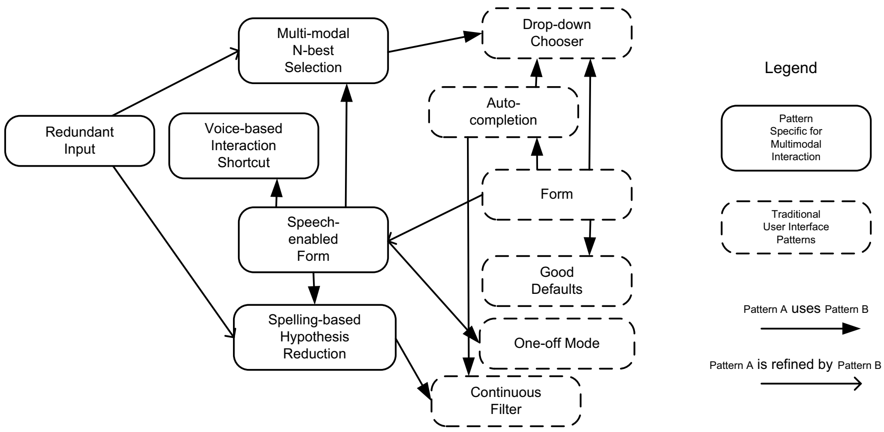

Fig. 4. Patterns for Multimodal Mobile Interaction

Following section describes the pattern *Speech-enabled Form* in some more detail. The remaining patterns can be found [36, 37].

Speech-enabled Form

Context
The user has to input structured data which can be mapped to some kind of form consisting of a set of atomic fields.

Devices such as PDAs do not provide a keyboard for comfortable string input. In other situations the device may support keyboard input but the user has only one hand available for interacting with the system.

This pattern is frequently used together with the patterns *Dropdown Chooser* [41] and *Autocompletion* [41]. For error handling and avoiding *Multi-modal N-best Selection* and *Spelling-based Hypothesis Reduction* can be used.

Problem
How to simplify string input in form filling applications?

Forces
- Selecting areas in 2D-space is accomplished comfortably with a pointing device but string input via pointing (with on-screen keyboards) is awkward.
- Values for some form items (academic degree, nationality etc.) are restricted and can be input using drop down choosers (combo boxes). But this may lead to screen clutter and additional navigation and scrolling.
- Speech recognition is very comfortable for selecting invisible items but the input of unconstrained text suffers from recognition errors.

Solution
Wherever possible determine acceptable values for each form field. Support value selection via *Dropdown Choosers* and, alternatively, via voice commands.

Let the user select the desired form field via pointing and input values via speech. The speech input complexity can be reduced, as only the vocabulary of the selected form item needs to be activated at the time.

In order to avoid that the speech recogniser interprets background noise as input, the recogniser should be activated only when the user is using speech input. One possibility is to activate the speech recogniser only while the user is holding down the pointing device over the desired entry field (cf. Tidwell's [41] pattern *Spring-loaded Mode*). Another possibility is to activate the speech recogniser for a certain time window after entry field selection (cf. Tidwell's [41] *One-off Mode*).

Consequences
- The user can comfortably combine pen input for selecting input fields with speech for value specification.
- Navigation and scrolling in drop down lists can be avoided.
- Constraining the voice recognition vocabulary according to the selected text field helps to avoid speech recognition errors.
- Speech recognition errors might occur anyway. In case of poor recognition performance all speed advantages might be lost due to the need of error corroboration.

Rationale
Users prefer speech to input descriptive data, or to select objects among large or invisible sets [20, 33].

In QuickSet, standard direct-manipulation was compared with the pen/voice multi-modal interface. Multi-modal interaction was significantly faster [12].

Known Uses
Mobile Systems such as Microsoft's MiPad [22] and IBM's Personal Speech Assistant [13] are good examples.

With MiPad the user can create e-mail messages via *Tap And Talk*. The user can select the addressee field and the speech recognition vocabulary is constrained to address book entries. If the user selects the subject or message field an unconstrained vocabulary is selected so that the user can input unconstrained text.

As a further example one could cite the QuickSet System [11].

The multi-modal facilities offered by X+V (XHTML and VoiceXML) and supported by the Opera Browser are heavily focussed on this *Speech-enabled Form* paradigm [23].

Related Patterns
This pattern is a multi-modal extension of *Form* as found in [40] and [38]. It is implemented using the pattern *Voice-based Interaction Shortcut* in the same way as *Form*s are implemented using patterns such as *Dropdown Chooser* and *Autocompletion*.

Tidwell's [41] patterns *Spring-loaded Mode* and *One-off Mode* can be used to control recogniser activation.

For error handling consider to use *Multi-modal N-Best-Selection* and *Spelling-based Hypothesis Reduction*.

5 Conclusion

This paper revealed the activities for mining patterns and creating a pattern language in emerging interaction paradigms of multimodal interaction. Modality properties and interaction constraints seem to give helpful advice in deciding which interaction technique should be used in which context. But for deeper design support more detailed guidelines or patterns are needed.

Patterns are identified both during top-down phases (based on multimodal interaction principles) and during bottom-up phases (based on pertinent use cases).

Recently, case studies involving empirical user tests on a multimodal email organiser both for desktop and mobile systems have been performed [35]. The results support the plausibility of this approach. In particular, the patterns *Voice-based Interaction Shortcut* and *Speech-enabled Form* were met with high user acceptance. This holds also for traditional interface patterns such as Tidwell's [41] *Autocompletion*. Tidwell's [41] *Spring-loaded Mode* or *One-off Mode* seem to be crucial for controlling recogniser activation.

6 References

1. Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., Angel, S.: *A Pattern Language.* Oxford University Press (1977)
2. Alexander, C.: *The Timeless Way of Building.* Oxford University Press (1979)
3. Almeida, L., Amdal, I., Beires, N., Boualem, M., Boves, L., den Os, E., Filoche, P., Gomes, R., Knudsen, J.E., Kvale, K., Rugelbak, J., Tallec, C., Warakagoda, N.: Implementing and evaluating a multimodal and multilingual tourist guide. In: *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems* (2002)
4. Aslan, I., Xu, F., Uszkoreit, H., Krüger, A., Steffen, J.: COMPASS2008: Multimodal, Multilingual and Crosslingual Interaction for Mobile Tourist Guide Applications. In: Maybury M. et al. (eds.) *Proceedings of Intelligent Technologies for Interactive Entertainment (INTETAIN)*, Springer, 3--12 (2005)
5. Benoît, C., Martin, J.C., Pelachaud, C., Schomaker, L., Suhm, B.: Audio-Visual and Multimodal Speech Systems. In Gibbon, D. (ed.) *Handbook of Standards and Resources for Spoken Language Systems - Supplement Volume* (1998)
6. Bernsen, N.O.: Multimodality in Language and Speech Systems – from theory to design support tool. *Lectures at the 7th European Summer School on Language and Speech Communication (ESSLSC)* (1999)
7. Borchers, J.: *A Pattern Approach to Interaction Design*. John Wiley & Sons, Inc. (2001)
8. Brown, W., Malveau, R., McCormick, H., Mowbray, T., Thomas, S.W.: The Software Patterns Criteria. web ressource: http://www.antipatterns.com/whatisapattern/ (1998)
9. Bürgy, C.: *An Interaction Constraints Model for Mobile and Wearable Computer-Aided Engineering Systems in Industrial Applications.* Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (2002)
10. Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., Stal, M.: *Pattern-orientierte Softwarearchitektur.* Addison-Wesley (1998)
11. Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., Clow, J.: QuickSet: multimodal interaction for distributed applications. In: *MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia*, ACM Press, 1997, 31--40 (1997)
12. Cohen, P., McGee, D., Clow, J.: The efficiency of multimodal interaction for a map-based task. In: *Proceedings of the sixth conference on Applied natural language processing*, Morgan Kaufmann Publishers Inc., 331--338 (2000)
13. Comerford, L., Frank, D., Gopalakrishnan, P., Gopinath, R., Sedivy, J.: The IBM Personal Speech Assistant. In: *Proc. of IEEE ICASSP'01*, 319--321 (2001)
14. Coutaz, J.I., Nigay, L., Salber, D., Blandford, A., May, J., Young, R.M.: Four Easy Pieces for Assessing the Usability of Multimodal Interaction: the CARE Properties. In: *Proceedings of INTERACT'95* (1995)
15. Dragičević, P.: *Un modèle d'interaction en entrée pour des systèmes interactifs multi-dispositifs hautement configurables*. Université de Nantes, école doctorale sciences et technologies de l'information et des matériaux. Diss. (2004)
16. Duarte, C., Carriço, L.: A conceptual framework for developing adaptive multimodal applications. In: *IUI'06: Proceedings of the 11th international conference on Intelligent user interfaces.* New York, NY, USA: ACM Press (2006)
17. Gabriel, D.: A Pattern Definition. web ressource: http://hillside.net/patterns/definition.html (2007)
18. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley (1995)

19. Godet-Bar, G., Dupuy-Chessa, S., Nigay, L.: Towards a System of Patterns for the Design of Multimodal Interfaces. In: *Proceedings of 6th International Conference on Computer-Aided Design of User Interfaces CADUI'2006*, Springer, 27--40 (2006)
20. Grasso, M.A., Ebert, D.S., Finin, T.W.: The integrality of speech in multimodal interfaces. In: *ACM Transactions on Computer-Human Interaction*. *5*(4) ACM Press, 303--325 (1998)
21. Hastie, H.W., Johnston, M., Ehlen, P.: CONTEXT-SENSITIVE HELP FOR MULTIMODAL DIALOGUE. In: ICMI '02: *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, IEEE Computer Society (2002)
22. Huang, X., Acero, A., Chelba, C., Deng, L., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, M., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Wang, K., Wang, Y.: MiPad: A Next Generation PDA Prototype. In: *ICSLP'00* (2000)
23. IBM: *Developing X+V Applications Using the Multimodal Toolkit and Browser*. Manual (2002)
24. Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker, S., Maloor, P.: MATCH: an architecture for multimodal dialogue systems. In: *Proc. of the 40th Annual Meeting on Association for Computational Linguistics*, A*CL '02*. 376--383 (2002)
25. Mahemoff, M.J., Johnston, L.J.: Usability Pattern Languages: the "Language" Aspect. In: Hirose, M. (ed.) *Human-Computer Interaction: Interact '01*. IOS Press for IFIP, 350--358 (2001)
26. Malaka, R., Haeussler, J., Aras, H.: SmartKom mobile: intelligent ubiquitous user interaction. In: *IUI '04: Proceedings of the 9th international conference on Intelligent user interface*, ACM Press, 310--312 (2004)
27. Marx, M., Schmandt, C.: Putting people first: specifying proper names in speech interfaces. In: *UIST '94: Proceedings of the 7th annual ACM symposium on User interface software and technology*, ACM Press, 29--37 (1994)
28. Milota, A.D.: Modality fusion for graphic design applications. In: *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, ACM Press, 167--174 (2004)
29. Neuss, R.: *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog*. Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Diss. (2001)
30. Niedermaier, F.B.: *Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug.* Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Diss. (2003)
31. Nigay, L., Coutaz, J.: A design space for multimodal systems: concurrent processing and data fusion. In: *Human Factors in Computing Systems, INTERCHI '93 Conference Proceedings*, ACM Press, 172--178 (1993)
32. Obrenović, Z., Abascal, J., Starčević, D.: Universal accessibility as a multimodal design issue. *Commun. ACM.* New York, NY, USA: ACM Press 50, 83-88 (2007)
33. Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D.: Designing the User Interface for Multimodal Speech and Pen-based Gesture Applications: State-of-the-Art Systems and Future Research Directions. In: *Human Computer Interaction. 15*(4), 263--322 (2000)
34. Ratzka, A., Wolff, C.: A Pattern-based Methodology for Multimodal Interaction Design. In: Sojka, P., Kopeček, I., Pala, K. (eds.) *Proc. of Text, Speech, and Dialogue, TSD'06*. Springer, 677--686 (2006)
35. Ratzka, A.: A Wizard-of-Oz Setting for Multimodal Interaction. An Approach to User-Based Elicitation of Design Patterns. In: Osswald, A., Stempfhuber, M., Wolff, C. (eds.)

Open Innovation. Proc. 10th International Symposium for Information Science, Universitätsverlag Konstanz, 159--170 (2007a)

36. Ratzka, A.: Design Patterns in the Context of Multi-modal Interaction. To appear in: *Proceedings of the 6th Nordic Conference on Pattern Languages of Programs 2007 VikingPLoP 2007* (2007b)
37. Ratzka, A.: Design Patterns for Robust and Accessible Multimodal Interaction. To appear in: *Proceedings of EuroPLoP 2008* (2008)
38. Sinnig, D., Gaffar, A., Reichart, D., Seffah, A., Forbrig, P.: Patterns in Model-Based Engineering. In: *CADUI'04*, 195--208 (2004)
39. Snelick, R., Indovina, M., Yen, J., Mink, A.: Multimodal biometrics: issues in design and testing. In: *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces*. ACM Press, 68--72 (2003)
40. Tidwell, J.: COMMON GROUND: *A Pattern Language for Human-Computer Interface Design*. web ressource: http://www.mit.edu/~jtidwell/interaction_patterns.html (1999)
41. Tidwell, J.: *Designing Interfaces: Patterns for Effective Interaction Design*. O'Reilly (2005)
42. van Welie, M., Trætteberg, H.: Interaction Patterns in User Interfaces. In: *Proceedings of the Seventh Pattern Languages of Programs Conference* (2000)
43. van Welie, M.: *Task-based User Interface Design*. Dutch Graduate School for Information and Knowledge Systems, Vrije Universiteit Amsterdam, Diss. (2001)
44. Zeng, Z., Tu, J., Liu, M., Zhang, T., Rizzolo, N., Zhang, Z., Huang, T.S., Roth, D., Levinson, S.: Bimodal HCI-related affect recognition. In: *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*. ACM Press, 137--143 (2004)