Identifying User Interface Patterns from Pertinent Multimodal Interaction Use Cases

Andreas Ratzka

Universität Regensburg

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. The topics of this work are not implementation centric patterns but rather user-task-near interface patterns. 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) are described, as well as more concrete use case specific patterns on the other hand. At first pattern candidates are derived from knowledge about how multimodality can enhance usability. At the same time literature is mined for real solutions as patterns are only valid as long as they describe proven solutions from the real world. Along with this, relationships between patterns are depicted in the context of these interaction use cases.

1 Introduction

The context of this work is usability engineering for multimodal interaction. Traditional approaches in this field focus on prototyping (cf. Niedermaier 2003, Dragičević 2004, Duarte & Carriço 2006) or decision support for requirements analysis and work reengineering (cf. Bernsen 1999, Bürgy 2002, Obrenović et al. 2007). The later stages in the usability engineering lifecycle i.e. design standards and detailed designs are only marginally covered by those decision support systems.

The idea of this work is to apply the concept of user interface patterns to the field of multi-modal interaction. A pattern is a rule connecting a common design problem with a proven solution and a description of the context indicating the conditions in which this pattern is applicable (Alexander et al. 1977, Alexander 1979, Brown et al. 1998, Gabriel 2007).

Patterns are based on well-proven design solutions of the real world. This seems impossible for the field of multimodal interaction which has not yet reached wide-spread market penetration. Nevertheless it can look back at almost thirty years of research during which several demonstration systems have been designed. Recurring problems that have lead to interesting solutions and were reused successfully in subsequent projects are a basis for identifying interface patterns from literature review (cf. Ratzka & Wolff 2006).

Successfully designing multimodal applications requires skills at several levels, such as among others software architecture (cf. the patterns in Buschmann et al. 1998), implementation techniques (cf. the patterns in Gamma et al. 1995), screen design, speech design and task modeling (cf. the patterns in van Welie & Trætteberg 2000; Borchers 2001).

This work focuses on the user-task-near type of user interface patterns. Even within this type of patterns one can distinguish different levels of granularity. This paper describes higher level patterns that are based on the general principles of multimodal interaction (patterns of multimodal combination and adaptation), as well as more concrete use case specific patterns.

The discussion of inter-pattern relationships is crucial as only this way a pattern collection forms a so called pattern language (cf. Mahemoff & Johnston 2001). Typical relationships are abstraction/refinement and usage/composition. This paper discusses relationships not only within the collection of multimodal interface patterns but also relationships to traditional user interface patterns such as those described in Tidwell (1999, 2005).

In terms of dual task scenarios, interactive maps and graphic design applications, traditional user interface patterns are discussed along with multimodal interaction techniques that alter traditional paradigms and thus lead to new, use-case-specific multimodal interface patterns. The identified patterns are shortly summarized in problem-solution tables. Detailed pattern descriptions can be found in Ratzka (2007, 2008).

2 Enhancing Usability via Multimodal Interaction

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 (adaptation) of appropriate interaction modalities at runtime.

2.1 Added Value via Modality Combination

Modalities are combined to minimize task interference, maximize information throughput, disambiguate distorted interaction channels, optimize saliency of notification messages and guarantee usability across different contexts of use, such as varying target users (think of disabilities), environment factors (noisy environments, privacy issues), and device characteristics (e.g. changing network bandwidth) which is critical for both public terminals and mobile interfaces.

Minimizing task interference

Multiple resource theory (Wickens 1980) and theory of working memory (Baddeley 1992) assume that there exist separated cognitive resources for each input and output modalities. Double task experiments have revealed that modality allocations that take this into account result in less interferences than if both tasks have to be performed using the same interaction channels (Wickens et al. 1984).

Maximizing information throughput

Assuming that multiple resource theory is right, multimodal presentations (spoken text plus graphic images) succeed in conveying more information to the user than unimodal ones (written text plus graphic images). Experiments on multimedia learning show that this way visual attention splits can be avoided (cf. Mayer & Moreno 1998). In the same way, the user is able to input complex information faster when he can select the most appropriate modality for each information atom (cf. Oviatt et al. 2004).

Mutual disambiguation of input information

When information is transmitted via several defective channels, it is more probable that at least some information can be conveyed at all than if only one channel is used. It is improbable that disturbances affect identical pieces of information on each channel. This principle is used among others in audiovisual speech recognition (cf. Benoît et al. 1998), person identification (cf. Snelick et al. 2003) and emotion recognition (cf. Zeng et al. 2004). This holds for the other communication direction, too. Multimodal perception of spoken language and associated lip movements improves human speech perception, i.e. the ratio of correctly recognized words by human (cf. Sumby & Pollack 1954, Summerfield 1979).

Assuring interaction across varying contexts of use

When it is not clear which interaction channels will work or are likely to be disturbed, several channels should be provided to the user to ensure interaction.

Optimizing saliency

Urgent information will be perceived with higher probability if it is presented via several channels simultaneously. Studies presented by Selcon & Taylor (1995) revealed that reaction time can be reduced by over 30 % the more redundant information channels are used. This issue is tackled by Tidwell's (1999) pattern *Important Message* which suggests important information (warnings) to be conveyed via several channels in parallel. Tidwell's collection contains also the "opposite" pattern *Status Display* which suggests status information to be presented in a non-disruptive way at a predictable place on the interaction surface.

Following problem-solution table exemplifies an exemplary subset of pattern candidates derived from the above listing.

Name	Problem	Solution
Audio- visual Workspace	How to minimize disruption in dual-task scenarios?	Present status information near to the visual attention area of the primary task. Present (urgent or at least immediately relevant) textual data such as route instructions via spoken language.
Audio- visual Presentation	How to convey complex information such as the functioning of technical systems or complex processes?	Combine several output media and modalities. Spatial data should be presented graphically. Concise conceptual data such as labels and keywords can be directly integrated into the graphic presentation. Extensive descriptions and explanations should be read out loudly.
Redundant Input	How to assure input when communication channels are distorted in an unforeseeable way?	Combine several interaction channels in order to make use of redundancy.

Table 1: Modality Combination Patterns

2.2 Added Value via Modality Adaptation

Systems that are used by different users subsequently, by individual users extensively (growing user expertise), in different or changing environments, or with changing degrees of service availability (due to changing network bandwidth) have to be adapted to these varying context factors. Adaptation can be done automatically (channel analysis, user modeling, etc.) or initiated by the user (changed behavior or explicit configuration changes). The following table outlines an exemplary pattern candidate derived from these issues.

Name	Problem	Solution
Multiple Ways of Input	How to ensure input although environment factors and target users are not known in advance?	Provide the user with the possibility to select his preferred interaction channel be it speech, typing or pointing.

Table 2: Modality Adaptation Patterns

3 Use-case Specific Solutions in Literature

Patterns never are inventions by their authors but always relate to successful examples of system design. This means that a pattern has to relate to at least three example systems which employed the described technique successfully before calling itself a pattern (Brown et al. 1998). This chapter provides the underpinnings for already discovered pattern candidates and identifies new ones from existing systems. This paper will focus on dual-task working environments, interactive maps and graphic design applications.

3.1 Dual Task Support Systems

Automotive and industrial applications frequently deal with situations where the user's eyes and hands are occupied with a primary task. Examples are driver assistance systems such as MoTiV-MMI (Bengler et al. 2000), CarMMI (Neuss 2001) or SIMBA (Salmen et al. 1999). Bürgy (2002) outlines several industrial applications for wearable computing which support similar dual task scenarios.

In these dual-task scenarios the pattern *Audio-visual Workspace* has to be used. Static information about the system's current state is displayed visually in a locally fixed area applying Tidwell's (1999) pattern *Status Display*. As Display area the car's centre console, head-up displaying techniques that project information onto the windscreen or, in mobile scenarios, head-mounted displays can be used.

In time information such as guidance directions are conveyed auditively in order not to distract the user's visual attention. Urgent data are displayed multi-modally using visual and auditive signals simultaneously as proposed in Tidwell's (1999) pattern *Important Message*.

3.2 Interactive Maps

Map-based systems are one of the widest spread applications of multi-modal interaction. Research projects such as *QuickSet* (Cohen et al. 1997), *SmartKom mobile* (Malaka et al. 2004), *MATCH* (Johnston et al. 2002) and *MUST* (Almeida et al. 2002) are examples.

These systems allow the user to combine (pen- or mouse-based) pointing input with spoken commands such as "zoom in here" or "grab this region". This recurring interaction technique is now called the pattern *Gesture-enhanced Natural Speech*. The most widespread use of this combination of speech and pointing are queries such as "show cheap Italian restaurants in this neighborhood <pen-gesture>" or "show me all psychiatrists <pen-gesture>". This technique is called the refined pattern *Multi-modal Spatial Search*.

Name	Problem	Solution
Gesture-enhanced Natural Speech	How to enable the user to quickly input composed commands consisting of several parameters?	Let the user interact via natural speech and provide pointing gestures simultaneously to specify loca- tions or interactive objects.
Multi-modal Spatial Search	Which interaction techniques allow the user to search efficiently for points of interests on an interactive map?	Combine pointing and speech input. Allow the user to select a region on the map via pointing and to input search terms via spoken language.

Table 3: Patterns found in Multimodal Interactive Maps

3.3 Graphic Design Applications

Computer Aided Design applications allow the user to create graphics using different kinds of tools. These applications are usually arranged based on the pattern *Canvas plus Palette* (Tidwell 2005) which means that the user can select the desired tool from the palette and apply this tool on the sketching surface, the canvas.

The tools on the palette usually switch modes and are used in combination with *Mode Cursor* (Welie & Trætteberg 2000). In this case a special cursor icon represents the selected tool and helps the user to predict system behavior.

Some tools, such as the pipette are usually used only once. Afterwards the user needs the tool he used beforehand. That's why it is appropriate to use Tidwell's (2005) pattern *One-off Mode* in these cases. In the case of *One-off Mode* the currently selected tool switches off itself after being used. *Spring-loaded Mode* (Tidwell 2005) can be used for special manipulation actions. As long as the user holds down a specific key, the behavior of the currently selected tool is modified.

The pattern *Canvas plus Palette*, however, forces the user to move the mouse cursor frequently between canvas and palette which might slow down interaction significantly. This is where Welie's (2001) pattern *Helping Hands* can help: The user uses one hand to input character shortcuts and the other one to perform graphical manipulations. This approach might reach its boundaries where shortcut characters are too cryptic to memories ore too cumbersome to input with the left hand alone.

Contextual Menus (Welie & Trætteberg 2000) are another approach to minimize mouse movements for action selection. But context menus might obscure parts of the central working surface without abolishing the need to remove the mouse cursor at all.

Multimodal design applications such as VoicePaint (Gourdol et al. 1992), S-tgif (Nishimoto et al. 1995) or Speak'n'Sketch (Sedivy & Johnson 2000) make use of spoken commands to avoid the necessity of repeated mouse movements between toolbars (palettes) and central working surface. This is a more convenient alternative to keyboard shortcuts and represents a versatile interaction pattern, called *Voice-based Interaction Shortcut* (cf. Ratzka 2007).

In order to control speech recognition activation, the user should be required to hold down the right mouse button while speaking. This is an application of Tidwell's (2005) pattern *Spring-loaded Mode*. More concretely, this new combination of *Canvas plus Palette*, *Spring-loaded Mode* and *Voice-based Interaction Shortcut* leads to the refining pattern *Speech-enabled Palette* (cf. Ratzka 2007).

Other graphic applications, especially those for mobile and tablet devices, build upon pen input and don't use palettes for lack of space but make use of another, sketch-based, paradigm (cf. Sezgin et al. 2001; Forbus et al. 2001). Rough sketches can be converted into geometric primitives via recognition techniques. In order to help the recognizer to find the desired geometric primitive, the user can input embellishing commands. Systems such as *TAPAGE* (Poirier et al. 1993), *QuickSet* (Cohen et al. 1997) or *DPD* (Milota 2004) allow the user to combine pen-based sketches and spoken embellishing commands. This special case

of Gesture-enhanced Natural Speech has been identified as new pattern and called Speech-enhanced Sketching.

Some graphic applications such as TAPAGE (Poirier et al. 1993) and *QuickSet* (Cohen et al. 1997) support special editing gestures, e.g. crossing out objects for deleting them. These gestures combine two information sources: the position where they are drawn (i.e. the object they are related to) and the shape they have, i.e. the meaning the convey (the command to be executed). This new pattern is now called *Location-sensitive Gesture*.

Name	Problem	Solution
Voice-based Inter- action Shortcut	How can the user quickly select an item from a large set?	To speed up interaction let the user select items via spoken language.
Speech-enabled Palette	How to enable the user to select tools from the palette without having to move the mouse between canvas and palette or the hand between mouse and keyboard?	Allow the user to select tools using speech input. Each tool on the palette should have a meaningful name which is being made obvious to the user to allow seamless learning.
Speech-enhanced Sketching	How to support graphic design with small pen-based devices?	Allow the user to input raw sketches with the pen and use spoken commands for embellishing.
Location-sensitive Gesture	How to enable the user to input easily and quickly commands consisting of selecting items and performing actions on them?	Let the user interact with the system as he would do with paper: draw meaningful symbols / pen gestures onto the object of interest – encircle items or cross them out, draw arrows etc.

Table 4: Patterns found in Multimodal Graphic Design Applications

4 Conclusion

This paper revealed an approach of mining user interface patterns from research literature on emerging paradigms of multimodal interaction. The use cases of dual task scenarios, map-based and design applications are examined in more detail and newly identified patterns are shortly outlined. In addition, the application of traditional user interface patterns in the above mentioned use cases is revealed along with relationships between individual (traditional and new) patterns.

The patterns outlined in this paper are part of a larger collection which forms an emerging pattern language for multimodal interaction (cf. Ratzka 2007, 2008). Recently, user tests with both desktop and mobile systems have been performed in order to provide empirical underpinnings. The results indicate the plausibility and user acceptance of the patterns adopted in the research prototypes.

In future, reusable building blocks (multimodal widgets) have to be defined which can be integrated into user interface builders and thus facilitate the implementation of the user interface patterns described in this work. These building blocks have to be open not only to the combination of traditional and speech interaction but also to emerging interaction styles that involve passive interaction modalities such as gaze input.

References

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., & Angel, S. (1977) *A Pattern Language*. Oxford University Press.

- Alexander, C. (1979). The Timeless Way of Building. Oxford University Press.
- 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. (2002). Implementing and evaluating a multimodal and multilingual tourist guide. In: Proc. of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems.
- Baddeley, A.D. (1992) Working memory. Science, 255, 556-559.
- Benoît, C., Martin, J. C., Pelachaud, C., Schomaker, L., & Suhm, B. (1998). Audio-Visual and Multi-modal Speech Systems. In: Gibbon, D. (ed.). *Handbook of Standards and Resources for Spoken Language Systems Supplement Volume*.
- Bengler K., Geutner P., & Steffens, F. (2000) "Eyes free Hands free" oder "Zeit der Stille". Ein Demonstrator zur multimodalen Bedienung im Fahrzeug. In: Gärtner, K. (ed.). *Multimodale Interaktion im Bereich der Fahrzeug- und Prozessführung*. DGLR-Report 200-02, 299-307.
- Bernsen, N.O. (1999). 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).
- Borchers, J. (2001). A Pattern Approach to Interaction Design. John Wiley & Sons, Inc.
- Brown W., Malveau, R., McCormick, H., Mowbray, T., & Thomas, S.W. (1998). The Software Patterns Criteria. web res.: http://www.antipatterns.com/whatisapattern/, checked on 2008-05-13.
- Bürgy, C. (2002). An Interaction Constraints Model for Mobile and Wearable Computer-Aided Engineering Systems in Industrial Applications. Dept. of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, Diss.
- Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., & Stal, M. (1998). *Pattern-orientierte Softwarearchitektur*. Addison-Wesley.
- Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., & Clow, J. (1997).
 QuickSet: multimodal interaction for distributed applications. In: MULTIMEDIA '97: Proc. of the fifth ACM international conference on Multimedia, ACM Press, 1997, 31-40.
- Dragičević, P. (2004). 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. Dissertation.
- Duarte, C. & Carriço, L. (2006). A conceptual framework for developing adaptive multimodal applications. In: *IUI'06: Proc. of the 11th international conference on Intelligent user interfaces.* New York, NY, USA: ACM Press.
- Forbus, K.D., Ferguson, R.W., & Usher, J.M. (2001). Towards a computational model of sketching. In: *IUI '01: Proc. of the 6th international conference on Intelligent user interfaces*, ACM Press, 77-83.
- Gabriel, D. (2007) A Pattern Definition. web res.: http://hillside.net/patterns/definition.html, checked on 2008-05-13.

- Gamma, E., Helm, R., Johnson R., & Vlissides, J. (1995). Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley.
- Gourdol, A., Nigay, L., Salber, D., & Coutaz, J. (1992), Two Case Studies of Software Architecture for Multimodal Interactive Systems: VoicePaint and a voice-enabled Graphical Notebook. In: Larson J. & Unger, C. (ed.). Proc. of IFIP TC3/WG2.7: Working Conference on Engineering for Human Computer Interaction, North Holland Publications, 271-284.
- Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker, S., & Maloor, P. (2002). MATCH: an architecture for multimodal dialogue systems. In: ACL '02: Proc. of the 40th Annual Meeting on Association for Computational Linguistics, ACL, 376-383.
- Mahemoff, M.J. & Johnston, L.J. (2001). Usability Pattern Languages: the "Language" Aspect. In: Hirose, M. (ed.) *Human-Computer Interaction: Interact '01*. IOS Press for IFIP, 350-358.
- Malaka, R., Haeussler, J., & Aras, H. (2004). SmartKom mobile: intelligent ubiquitous user interaction. In: *IUI '04: Proc. of the 9th international conference on Intelligent user interfaces*, ACM Press, 310-312.
- Mayer, R.E. & Moreno, R. (1998). A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory. In: *Journal of Educational Psychology*, 90(2), 312-320.
- Milota, A.D. (2004). Modality fusion for graphic design applications. In: *ICMI '04: Proc. of the 6th international conference on Multimodal interfaces*, ACM Press, 167-174.
- Neuss, R. (2001). Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog. Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Diss.
- Niedermaier, F. B. (2003). Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug. Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Diss.
- Nishimoto, T., Shida, N., Kobayashi, T., & Shirai, K. (1995). Improving Human Interface in Drawing Tool Using Speech. Proc. of 4th IEEE International Workshop on Robot and Human Communication, ROMAN'95, 107-112.
- Obrenović, Z., Abascal, J., & Starčević, D. (2007). Universal accessibility as a multimodal design issue. *Commun. ACM*, 50(5). New York, NY, USA: ACM Press, 83-88.
- Oviatt, S., Coulston, R., & Lunsford, R. (2004). When do we interact multimodally?: cognitive load and multimodal communication patterns. In: ICMI '04: Proc. of the 6th international conference on Multimodal interfaces, ACM Press, 129-136
- Poirier, F., Julia, L., Rossignol, S., & Faure, C. (1993). TAPAGE: édition de tableaux sur ordinateur à stylo vers une désignation naturelle. In: *Proc. IHM'93*.
- Ratzka, A. & Wolff, C. (2006). A Pattern-based Methodology for Multimiodal Interaction Design. In: Sojka, P. et al. (ed.). *Proc. of Text, Speech, and Dialogue, TSD'06*, Springer, 677-686.
- Ratzka, A. (2007). Design Patterns in the Context of Multi-modal Interaction. To appear in: *Proc. of the 6th Nordic Conference on Pattern Languages of Programs 2007 VikingPLoP 2007.*
- Ratzka, A. (2008). Patterns for Robust and Flexible Multi-modal Interaction. To appear in:. Proc. of EuroPLoP 2008.

Salmen, A., Grossmann, P., Hitzenberger, L., & Creutzburg, U. (1999). Dialog Systems in Traffic Environment. In: Proc. of ESCA: Tutorial and Research Workshop on Interactive Dialogue in Multi-Modal Systems.

- Sedivy J. & Johnson, H. (2000). Multimodal tool support for creative tasks in the visual arts. Knowledge-Based Systems, 13(7-8), 441-450.
- Selcon, S.J. & Taylor, R.M. (1995). Integrating Multiple Information Sources: Using Redundancy in the Design of Warnings. In. *Ergonomics* 38(11), 2362-2370.
- Sezgin, T.M., Stahovich, T., & Davis, R. (2001). Sketch based interfaces: early processing for sketch understanding. In: *Proc. of the 2001 workshop on Perceptive user interfaces*, ACM Press, 1-8.
- Snelick, R., Indovina, M., Yen, J., & Mink, A. (2003). Multimodal biometrics: issues in design and testing. In: Proc. of the 5th international conference on Multimodal interfaces, ACM Press, 68-72.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. In: *Journal of the Acoustical Society of America*, 26, 212-215.
- Summerfield, A.Q. (1979). Use of visual information for phonetic perception. In: *Phonetica, 36,* 314-331.
- Tidwell, J. (1999). COMMON GROUND: A Pattern Language for Human-Computer Interface Design, web res.: http://www.mit.edu/~itidwell/common ground.html, checked on 2008-05-10
- Tidwell, J. (2005). Designing Interfaces: Patterns for Effective Interaction Design. O'Reilly.
- van Welie, M. & Trætteberg, H. (2000). Interaction Patterns in User Interfaces. In: *Proc. of the Seventh Pattern Languages of Programs Conference*.
- van Welie, M. (2001). *Task-based User Interface Design*. Dutch Graduate School for Information and Knowledge Systems, Vrije Universiteit Amsterdam, Diss.
- Wickens, C. (1980). The Structure of Attentional Resources. In: Nickerson, R.S. (ed.). *Attention and Performance VIII. Lawrence Erlbaum*, 239-257.
- Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of S-R-C compatibility with spatial and verbal tasks: The role of display-control interfacing. In: *Human Factors*, 26, 533-534.
- Zeng, Z., Tu, J., Liu, M., Zhang, T., Rizzolo, N., Zhang, Z., Huang, T.S., Roth, D., & Levinson, S. (2004). Bimodal HCI-related affect recognition. In: ICMI '04: Proc. of the 6th international conference on Multimodal interfaces, ACM Press, 137-143.

Contact Information

Andreas Ratzka Lehrstuhl für Informationswissenschaft Universität Regensburg D-93040 Regensburg

Tel.: +49 941 943 3464

E-Mail: Andreas.Ratzka@sprachlit.uni-regensburg.de