

MULTIFACE: Multimodal Content Adaptations for Heterogeneous Devices

Songsak Channarukul Susan W. McRoy Syed S. Ali
Natural Language and Knowledge Representation Research Group
Department of Computer Science
University of Wisconsin-Milwaukee
{songsak,mcroy,syali}@tigger.cs.uwm.edu

ABSTRACT

We are interested in applying and extending existing frameworks for combining output modalities for adaptations of multimodal content on heterogeneous devices based on user and device models. In this paper, we present **Multiface**, a multimodal dialog system that allows users to interact using different devices such as desktop computers, PDAs, and mobile phones. The presented content and its modality will be customized to individual users and the device they are using.

Categories and Subject Descriptors: H.5.2 [User Interfaces]: User-centered design

General Terms: Design, Human Factors.

Keywords: Multimodal output, device-centered adaptation, user-centered adaptation, dialog system.

1. INTRODUCTION

Dialog systems that adapt appropriately to different user needs and preferences have been shown to achieve higher levels of user satisfaction [9]. However, it is also important that dialog systems be able to adapt to the user's computing environment, because people can access computer systems using different devices.

Existing research on device-centered adaptations ranges from low-level adaptations such as conversion of multimedia objects [11] (*e.g.*, video to images, audio to text, image size reduction) to higher-level adaptations based on multimedia document models [3]. In our work, we are investigating an approach that does not use either low-level adaptations or multimedia document models. This is because low-level adaptations are time-consuming, whereas multimedia document models require multiple copies of the same document that are pre-authored for different device types. Instead, we are interested in applying and extending the frameworks for combining output modalities [7, 10] to multimodal dialog systems that target heterogeneous devices.

This paper describes **Multiface**, a multimodal dialog system that adapts its content and modalities based on both users and device types.

2. ADAPTING MULTIMODAL CONTENT

Multiface customizes its content and determines the appropriate modality based on individual users and their computing environment. For example, if the user is using a desktop computer with high bandwidth Internet, **Multiface** can opt for heavy uses of video and audio. However, if the bandwidth is low, a video might be substituted by pictures with captions. In some situations, a handheld computer is more appropriate for the user. For example, in the tutoring domain, users might want to review their past lessons while commuting, or need a summary of some procedures while practicing in the real-world. The device and the intended purpose of accessing the system will also affect the style of the presented information. Presentation styles might range from a declarative format (textual and lengthy) to a procedural format (more precise and imperative). Both user and device models are taken into consideration to achieve well-customized content and presentation.

2.1 Content Selection

Multiface's approach to adaptive selection of content is based upon the approach used in **Equuleus** [8]. **Equuleus** is a system for interactive, adaptive, document presentation. It adapts the content to individual users, based on their level of expertise, what sections they have already read, and how they have performed on related tests. **Equuleus** achieves this by using an annotated document that uses XML tags to indicate boundaries of sections of the document and semantic (rhetorical) links among those sections. The semantic markup is based on Rhetorical Structure Theory [6].

2.2 Modality Selection

Content modalities in **Equuleus** are limited to text and image. There is no mechanism that allows adaptations of content modalities. To do this, **Multiface** enhances the document by tagging existing content units with a modality such as *text* or *image*. Additional content modalities such as *audio*, *video*, and *summary* are added to the annotated document to allow multimodal content adaptations.

Multiface employs modality relations that are represented as rules in its planner (a JAM agent [5]). These modal-

ity relations are based on the TYCOON framework [7]. In this framework, there are six cooperation types between input modalities such as *equivalence* (each modality equally convey the same information) and *complementarity* (one modality complements another modality). These cooperation types define functional relationships between modalities. For example, an audio and its textual counterpart are considered *equivalent*. An image and a caption are considered *complementary*. In the case of audio and text, they might be presented at the same time if both modalities are feasible for the user given the current computing environment. However, if it is not appropriate to present an audio (e.g., when a user is in a loud environment, when a device does not have speakers), presenting only text is sufficient.

Beside modality relations, Multiface also takes into account a device model that describes the current device of a user. Multiface uses an object-oriented approach to device modeling where devices that share the same properties are grouped into a finite number of types. Each device type captures device properties at a more abstract level by classifying devices into primary types (e.g., desktop computer, handheld computer, and mobile phone), defining what modalities that are *allowed* and *preferred* for each device type, and specifying other devices that belong to the existing device type (i.e., share some modality properties) but have different properties in some modalities. By *allowed*, we mean the device is capable of presenting a particular modality. *preferred* means the device is good at presenting such a modality. This approach requires less effort to build a device model and also allows specification of more devices when necessary. Automated derivation from existing device models (e.g., CC/PP [2]) to our device model is under investigation.

3. MULTIMODAL OUTPUT GENERATION

Upon selection of appropriate content and modalities, Multiface needs to produce an output that can be displayed on various devices. Multiface employs the DOGHED system [4] to generate output in the Synchronized Multimedia Integration Language (SMIL) [1]. An example of Multiface's output is shown in Figure 1.

4. REFERENCES

- [1] <http://www.w3.org/TR/smil20>, Synchronized Multimedia Integration Language (SMIL 2.0). W3C, 2001.
- [2] <http://www.w3.org/TR/CCPP-struct-vocab>, Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0. W3C, 2004.
- [3] S. Boll, W. Klas, and U. Westermann. A comparison of multimedia document models concerning advanced requirements. Technical Report 99-01, Ulmer Informatik-Berichte, University of Ulm, Germany, 1999.
- [4] S. Channarukul, S. W. McRoy, and S. S. Ali. DOGHED: A Template-Based Generator for Multimodal Dialog Systems Targeting Heterogeneous Devices. In *Companion Volume to the Proceedings of HLT-NAACL 2003*, pages 5–6, Canada, May 2003.
- [5] M. J. Huber. JAM Agents in a Nutshell. Technical report, Intelligent Reasoning Systems, November 2001.



Figure 1: An Example of Multiface's output

- [6] W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. In *TEXT*, volume 8:3, pages 243–281. 1988.
- [7] J. Martin. Six primitive types of cooperation for observing, evaluating, and specifying cooperations. In *Proceedings of American Association for Artificial Intelligence Conference*, pages 35–51. Springer-Verlag, 1999.
- [8] S. W. McRoy, S. S. Ali, and N. Nalamlieng. Equuleus: Presentation from Legacy Documents. In *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*, pages 589–596, November 2003.
- [9] A. Stent, M. Walker, S. Whittaker, and P. Maloor. User-Tailored Generation for Spoken Dialogue: An Experiment. In *Proceedings of ICSLP 2002*, pages 1281–84, September 2002.
- [10] F. Vernier and L. Nigay. A framework for the combination and characterization of output modalities. In *Proceedings of DSV-IS2000, Lecture Notes in Computer Science*, pages 32–48. Springer-Verlag, 2000.
- [11] A. Vetro and H. Sun. Media conversions to support mobile users. In *Proceedings of The IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 607–612, 2001.