Chapter 9

QUALITY OF EXPERIENCING MULTI-MODAL INTERACTION

Benjamin Weiss, Sebastian Möller, Ina Wechsung and Christine Kühnel
Quality & Usability Lab, Deutsche Telekom Laboratories
TU Berlin, Germany
{BWeiss,Sebastian.Moeller,Ina.Wechsung,Christine.Kuehnel}@telekom.de

Abstract In this chapter, we discuss the contributions of different modalities to the overall quality of multi-modal interaction. After reviewing some common systematics and findings concerning multi-modality, we present experimental results from several multi-modal scenarios, involving different (human-to-human and human-to-machine) interaction paradigms, different degrees of interactivity, and different (speech, audio, video, touch, gesture) modalities. The results show that the impact of each modality on overall quality in interaction depends heavily on the scenario and degree of interactivity. Complementary modalities are not considered in this paper, but the models presented allow predicting overall system quality on the basis of individual modality ratings with an appropriate accuracy. These models still have to be validated in order to be used as tools for system developers estimating whether adding modalities will have an impact on the quality experienced by the user.

Keywords: Usability; User experience; Perceived quality; Multi-modal integration.

1. Introduction

Multi-modal dialog systems appear to offer better interaction experience, as multi-modality seems to have fundamental advantages over unimodal interaction. However, there are few matching examples beyond the standard “put-that-there” scenario. Much more often, simply providing alternative input or output modalities resulting in sequential multi-modality seems to be the state-of-the-art. The question is what constitutes a “good” interaction, i.e. what aspects contribute to the user having a good or bad impression of the system she has been using. This is commonly understood by the term “Quality of Experience”, QoE.
In this chapter, we will summarize major results concerning multi-modality at first, and then provide a common ground on what Quality of Experience really means. We will then present experimental results from different interaction scenarios: Audio-visual transmission systems (like IP-based television or audio-visual telephony), interactions with Embodied Conversational Agents (ECAs), as well as interactions with different non-embodied multi-modal dialogue systems providing speech, touch and motion input capabilities. For each scenario, algorithmic models are presented which quantify the impact of each modality on the overall system quality, as it is perceived subjectively by the user. The goodness of the models are described in term of Pearson’s correlation $R$ between the models’ estimates and the real data obtained, as well as the root mean squared prediction error (RMSE). We conclude by identifying some research questions which should be answered in order to fully support the design and evaluation of multi-modal dialog applications.

2. Advantages of Systems Providing Multi-Modal Interaction

One major assumption concerning human-computer interfaces is that the interaction is significantly facilitated by providing multiple input modalities and by presenting information over different output channels. From a usability point of view – i.e. discounting hedonic aspects like appearance and style of the interface or the possibility to express the user’s identity with a given product – a multiple of possible input modalities can increase the recognition rate by fusing different input modalities (e.g. on the signal level) and it allows people to use those modalities most adequate in their specific situation, mood and capability (López-Cózar Delgado and Araki, 2005; Oviatt, 2004). For example, touch may be favoured in noisy or public environments, speech for the task of selecting objects in longer lists, typing for editing text and pointing gestures to refer to spatial information. Concerning the system output, multiple modalities allow for selecting the most appropriate way to present a specific piece of information (e.g. Graphical User Interfaces for lists, Embodied Conversational Agents for emotions, auditory icons for alarms, short vibrations for positive feedback). Another benefit is the possibility to present information redundantly to increase salience.

Furthermore, there seem to be cognitive advantages for multi-modal interfaces. Redundant and complementary information may distribute the use of cognitive resources and thus make processing faster and less demanding. With the theory of multiple resources (Wickens, 1999) for example, the tasks of speaking and gesturing or hearing and watching use different resources that in principle should not interfere with each other. As a result, users seem to prefer multi-modal interaction, especially, when the cognitive load increases