

You Cannot Be Serious: Towards Objective Criteria for Defining Multimodal Interaction

BLUEBOOK NOTE (Inspired by discussions at [the MOG workshop](#))

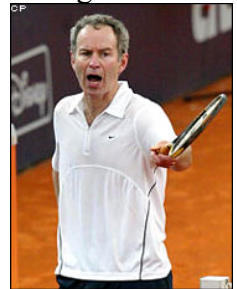
Zeljko Obrenovic

Created: Sept 7, 2007

Last updated: Sept 7, 2007

Introduction

Tennis player John McEnroe made famous the phrase "you cannot be serious," by shouting it after some umpires' calls during his matches [[McEnroe](#)]. Similar phrases can also be heard in discussions that try to define what multimodal interactions is, i.e. when the interaction is multimodal, and when it is not. Some people treat features such as color or text as modalities, some just laugh at it, insisting that the interaction is multimodal if it combines two different perceptual channels such as visual and audio.



We wanted to show that both sides are right, but that they are using different measures and granularity to define the term. In this paper, we introduce new more objective criteria for defining multimodal interaction, and distinguishing it from other forms of HCI.

Existing "Definitions" of Multimodal Interactions

In computer sciences, the meaning of the term "modality" is ambiguous. In human-computer interaction, the term usually refers to the human senses—vision, hearing, touch, smell, and taste—but many researchers distinguish between computing modalities and the sensory modalities of psychology.¹

Sharon Oviatt offered a more practical definition, saying that multimodal systems coordinate the processing of combined natural input modalities—such as speech, touch, hand gestures, eye gaze, and head and body movements—with multimedia system output.² Matthew Turk and George Robertson further refined the difference between multimedia and multimodal systems, saying that multimedia research focuses on the media, while multimodal research focuses on human perceptual channels.³ They added that multimodal output uses different modalities, such as visual display, audio, and tactile feedback, to engage human perceptual, cognitive, and communication skills in understanding what is being presented. Multimodal interaction systems can use various modalities independently, simultaneously, or by tightly coupling them

In his [Modality Theory](#) Bernsen noted correctly that when people talk about modalities, they immediately talk about multimodal interaction. However, what is unimodal interaction is never precisely defined. Bernsen propose its taxonomy of pure unimodal items that are combined to create multimodal interaction. But the ground for his classification is not defined precisely.

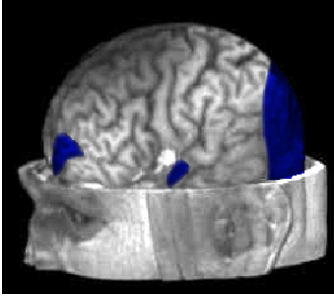
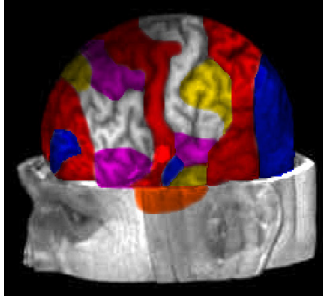
New Criteria for Definition of Multimodal vs. Unimodal Interaction

We propose two (interconnected) criteria for defining multimodal interaction and distinguishing it from unimodal interaction: brain regions active during the interaction, and human functionalities used during the interaction. The two criteria are interconnected, i.e. human functionalities can be expressed in terms of brain regions they require. However, using standardized vocabulary, such as ICF, enables relation of multimodal research with other research about human functionalities, disabilities and health.

Brain Regions Activation As Criteria

- DEF: Multimodal interaction is interaction which simultaneously involves different regions of human brain. Therefore, when the interaction will be multimodal, and when unimodal, depends how you define these regions. In simple case, we can define it according to cortex: visual, motor, audio, olfactory... However, it also makes sense to define subregions in these cortex, at least in visual cortex, which occupies 25% of cortex, and is bigger than audio and motor cortex together. Moreover, there are more and more evidence that information within one sensory channel also follows similar integration mechanisms as multisensory integration. Our understanding of multisensory integration has advanced because of [recent functional neuroimaging studies](#). Beyond identifying existence of multisensory integration mechanisms, these studies also indicate existence of a neural mechanism for integrating disparate representations within individual sensory modalities, such as representations of visual form, color, and visual motion.
- We can also argue that the goal of multimodal interaction is to maximize the brain usage, and in this way we can even quantify this. Hypothesis is that multimodal combination that activates more brain power is better. We can also identify if modalities have conflicting requests for brain power.

Examples...

	
A brain activation during today's dominant unimodal input and unimodal output (where visual cortex is treated as atomic region, and visual display is only output, and mouse and keyboard are only input devices).	A brain activation during multimodal interaction

Human Functionalities As Criteria

- DEF: Other possible criteria is defining a multimodal interaction as an interaction that simultaneously involves complementary human functionalities (based on [ICF](#) classification). Again, ICF is multi-level taxonomy, so depending on which level we compare, answer to the question what constitutes the multimodal/unimodal interaction can be different.
- Quantifying the level of human functionalities required by some multimodal combination

Examples...

Advantages of New Criteria

- Objective definition of multimodality vs. unimodality
- Relation with Accessibility research
- Identifying conflicting requirements of modalities
- Metric for defining the measure of Multimodality
 - How much of the brain power is used
- Point to area in brain that can be exploited more