FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Quantitative Analysis of Gestures in Relation to Activity Categories when Working with the ColorTable

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

im Rahmen des Studiums

Medieninformatik

eingereicht von

Gudrun Wimmer, Bakk.techn.
Matrikelnummer 0009472

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung
Betreuer/in: Univ.Prof. Dr.phil. Ina Wagner

Wien, 03.11.2011

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

Erklärung zur Verfassung der Arbeit

Gudrun Wimmer, Gottschalkgasse 1/23, 1110 Wien

„Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen – die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe."

Wien, 03.11.2011

Abstract

The ColorTable is an urban renewal application that enables a diverse group of stakeholders to share their ideas about an urban planning issue. The system is a tangible user interface with which the participants can develop interventions for an urban site together and translate them into virtual mixed reality scenarios.

When working with the ColorTable the participants engage in various activity categories. This thesis is a quantitative analysis that examines different aspects of behavior during each of these categories, with a special focus on the role of gestures.

The analysis identifies very clear patterns of behavior for each category combining the behavioral aspects mentioned above. Specifically, a lot of insight can be gained by examining the focus of pointing gestures. This leads to the conclusion that the ColorTable would greatly benefit from developing a way to incorporate pointing into the system.

Zusammenfassung

Der ColorTable ist eine spezielle Anwendung für die Stadtentwicklung, die es einer Gruppe von unterschiedlich qualifizierten Akteuren ermöglicht, Stadtplanungskonzepte zu entwickeln. Das System besteht aus einer greifbaren Benutzerschnittstelle (engl. Tangible User Interface), mit der die Beteiligten gemeinsam ihre Interventionen erarbeiten und als eine virtuelle Mixed-Reality-Szene darstellen können.

Bei der Arbeit mit dem ColorTable durchlaufen die Teilnehmer verschiedene Kategorien von Aktivitäten. Die vorliegende Diplomarbeit untersucht in einer quantitativen Analyse verschiedene Facetten von Verhalten in den jeweiligen Kategorien mit einem speziellen Schwerpunkt auf Gesten.

Die Analyse der Verhaltensweise ermöglichte die Identifizierung von eindeutigen Verhaltensmustern in Bezug auf die Kategorien. Vor allem die Betrachtung der Zeigegesten lieferte weitgreifende Erkenntnisse. Daraus ergibt sich die Schlussfolgerung, dass der ColorTable davon profitieren würde, Zeigegesten in das System zu integrieren.

Acknowledgments

First of all I would like to express my gratitude to my supervisor Prof. Ina Wagner for her guidance and invaluable suggestions. My thanks also go to the IPCity Research Team for making me feel welcome and sharing their insights.

I would like to thank my family and friends for their patience and support. I am especially grateful to my partner, Joachim Bickel, for the valuable input and technical assistance.

Contents

1 Introduction

1.1 Motivation

The *IPCity – Integrated Project on Interaction and Presence in Urban Environments* was funded by the European Union and consisted of a number of independent projects. Each of these projects developed a novel technological system that would open up new perspectives on urban environments for users.

One of these projects is the *ColorTable,* which was developed at the Institute for Design & Assessment of Technology – Multidisciplinary Design Group at the Vienna University of Technology. The system is a platform for multi-disciplinary urban planning teams to develop solutions for urban sites by creating mixed reality scenes with a tangible user interface.

To conclude the project the design team conducted an in-depth qualitative analysis of the ColorTable experience. The report for IPCity describes the final stage of an evaluation-feedback-redesign process that guided the creation of the ColorTable and the evaluation of its usage. It includes the development and technical setup of the last prototype, a description of two participatory workshops and the qualitative analysis itself, which was structured around eight research questions and four different activity categories. The design team uncovered a number of findings that led to the definition of design guidelines.

Among other insights, the results showed that gestures and especially pointing gestures played a major role in working with the ColorTable and understanding the virtual scenes. Furthermore, gestures are in general an important part of communication.

The aim of this master's thesis is to expand on these findings by performing a quantitative analysis of the behavior around the ColorTable. At the center of the analysis are gestures, which are allocated according to their characteristics to one of six different categories: iconic gestures, pointing, dual pointing, tracing, encircling, and beats.

But solely looking at gestures is not enough. They have to be viewed in the context in which they occurred. The qualitative analysis plays a major role along with the other activities around the ColorTable in contextualizing them. This includes how the participants use the features of the ColorTable to make changes and explore the MR scene. To complete the picture of the interaction in the MR-Tent the spatial distribution of the participants and the group dynamics are also taken into consideration. By uncovering new patterns of behavior it should be possible to deduce recommendations for improvements.

Because gestures are more complex and require more background information than the other behaviors examined, the thesis also includes a literature review of and some conceptual work on their characteristics, which provides the basis for correctly identifying gestures and assigning them to the different categories.

1.2 Thesis Roadmap

This master's thesis presents a quantitative analysis that examines the activities of the participants within the environment of the ColorTable. The spectrum of the analysis covers the usage of application functions, spatial distribution of the participants, and has a special focus on gestures. The framework for analysis is derived from the findings of the qualitative analysis.

Chapter 2 is an introduction to the MR-Tent and the ColorTable. This includes the design process that shaped the system, an overview of the integrated external software components, and the functionality and haptic features of the ColorTable.

Working with gestures requires a detailed theoretical look on what constitutes a gesture and how they can be categorized. A definition of gestures is delineated in Chapter 3. This chapter also examines the link between gesture and speech. At this point it is established what a gesture is and Chapter 4 provides the tools to distinguish gestures from other nonverbal behavior and concludes with the categories for the analysis. One of these categories is deictic gestures or pointing and because of their importance they are further discussed and refined into subcategories in Chapter 5.

The data for the analysis was collected during a participatory workshop in France. Chapter 6 gives an overview of the course of events at the workshop and the methods of data collection are outlined in Chapter 7.

The methodical framework and the research questions at the core of the qualitative analysis are summarized in Chapter 8 along with the findings that pertain to gestures.

This informs the construction of the framework for the quantitative analysis and culminates in the quantitative analysis itself in Chapter 9. The usage of gestures is analyzed in the context of activities and use of space around the ColorTable and related to the activity categories that were the cornerstones of the qualitative analysis.

The conclusions from the behavioral patterns that manifested themselves in Chapter 9 and the resulting design recommendations are presented in the last chapter alongside areas of future work.

2 The MR-Tent

2.1 Design Process and Technologies

The MR-Tent prototype was developed at the Institute for Design & Assessment of Technology – Multidisciplinary Design Group at the Vienna University of Technology. The main motivation behind the MR-Tent was to create an environment in which urban planners, inhabitants, city officials, and other stakeholders can develop and discuss urban interventions in collaboration. The development was based on a participatory design process and a multi-disciplinary team that included members from the fields of computer science, social science, art, design, and urban planning.

The MR-Tent underwent a number of development-evaluation-redesign cycles. Each of these cycles was built around a *participatory workshop*. During these workshops the participants worked on existing urban planning projects, which gave them the opportunity to explore the possibilities and limitations of the different prototypes. These sessions were documented with video recordings, digital cameras, and screenshots of the created virtual scenes. The collected data was then used by the design team to analyze the interactions and discussions of the participants. The insights gleaned from these analyses were applied to enhance the existing prototype or construct a new one with the requirements of the following workshop in mind. For more detailed discussions of the design process see [1].

The resulting MR-Tent combines a number of preexisting components, for example Color Tracking [2], UrbanSketcher [3] or Hypermedia Database [4], that have been modified and considerably extended to create a smooth workflow for the participants. The ColorTable is the technological core of the MR-Tent.

Because the entire setup is supposed to travel to different urban planning sites, the tent was developed in conjunction with the ColorTable during the design process. The MR-Tent, which houses the equipment on site, creates a meeting space for the participants in which they can safely share their ideas, while at the same time it can also be opened up to the surrounding area, thus creating a direct connection to the site of the urban planning project. Figure 1 was taken on site during the workshop in Pontoise.

Figure 1: The MR-Tent on site in Pontoise [5]

2.2 Components and Features

The video material that is the basis for the analysis to follow was taken during a workshop with the final prototype of the ColorTable, which is described in [1] and [5]. These papers are the basis for the summary of the components and features of the final prototype.

The ColorTable allows users to share their ideas by placing content into a virtual scene. The users can connect different urban elements, for example buildings, walkways, roads, or ground textures, to colored tokens of varying shapes. In order to arrange the elements in the scene, the tokens are placed on a physical map of the project site. On a vertical projection the interventions are displayed in real time by rendering them in 3D

into the virtual scene. The technical setup and a view inside the tent are shown in Figure 2 and Figure 3, respectively.

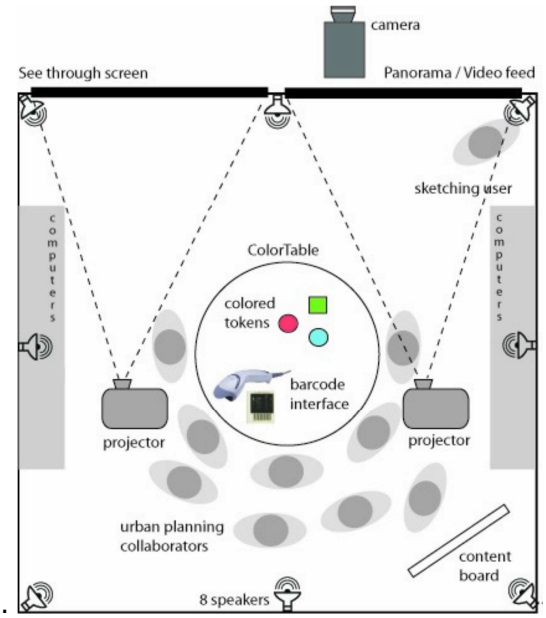


Figure 2: The technical setup within the MR-Tent [6]

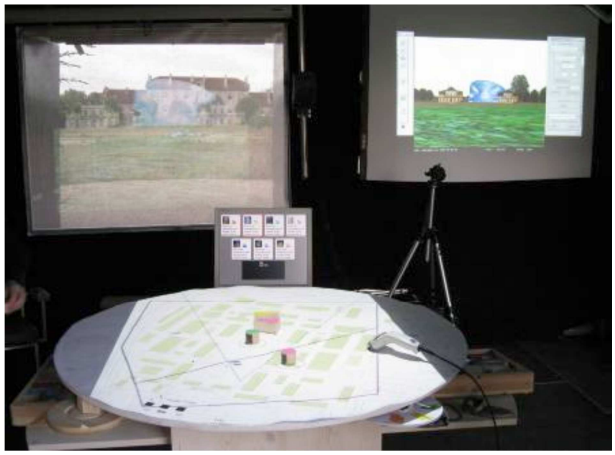


Figure 3: View inside the MR-Tent [6]

2.2.1 The ColorTable

From boardrooms to dining rooms, tables are gathering places for people and centers of conversations. The round shape of the ColorTable fosters a nonhierarchical environment for collaboration. The surface of the table is the staging area for the planning of urban interventions. The users can choose between maps of different region and scales and place them on the table. The maps are identified with barcodes and also include barcodes which correspond to viewpoints of the virtual panoramic scenes.

Barcode Reader

The viewpoints are selected with the *barcode reader* that is attached to the table. The barcode reader can also be used to select different functions. For example, the *rotating wheel* can either be used to zoom in or out of the panorama or to rotate along the panorama. The barcodes for the miscellaneous functions are accessible on a separate cardboard. For example, the *freeze-function* saves the current scenario. The users can than remove the tokens from the table and add additional content to the saved scene. The saved scenes remain accessible for browsing through later on.

Top-Down View

A projection from the top unto the table augments the maps and tokens with further information. Digital lines and dots show the course of roads and flows. Small thumbnails represent the content assigned to a token. Furthermore, feedback is provided on whether a token was successfully detected by the tracking software.

Configuration Board

The configuration board uses RFID technology to assign content to the tokens. The content is represented through physical *content cards*. They are small rectangular cardboards with a thumbnail representation of their digital content and the shape of the token they can be assigned to. In order to provide the users with an overview of the available contents, the cards are arranged on the *content board*. Users can gather in

front of the content board and discuss their selection of content together. The chosen cards are placed on the colored fields of the configuration board, thus connecting them with the token of the same color and shape that is defined by the digital content.

A special subset of content cards is used to change attributes of the digital content. By placing these cards on one of the colored fields, the user can manipulate the size, offset (the distance from the ground), spacing (of objects placed along a line), and color overlay of the 3D objects and billboards assigned to the field. (A detailed description of the possible content is given below in the part that deals with the tokens.) For example, the size, offset, and spacing have different cards for increasing and decreasing their value. While these cards are placed on the field the values increase or decrease incrementally. When the desired value is reached the card is removed.

Information Area

Above the configuration board, on a piece of paper that overlays a small part of the map, is a projection of the vital information of the object that was last manipulated. The information can also be requested by placing a special content request card on one of the colored fields.

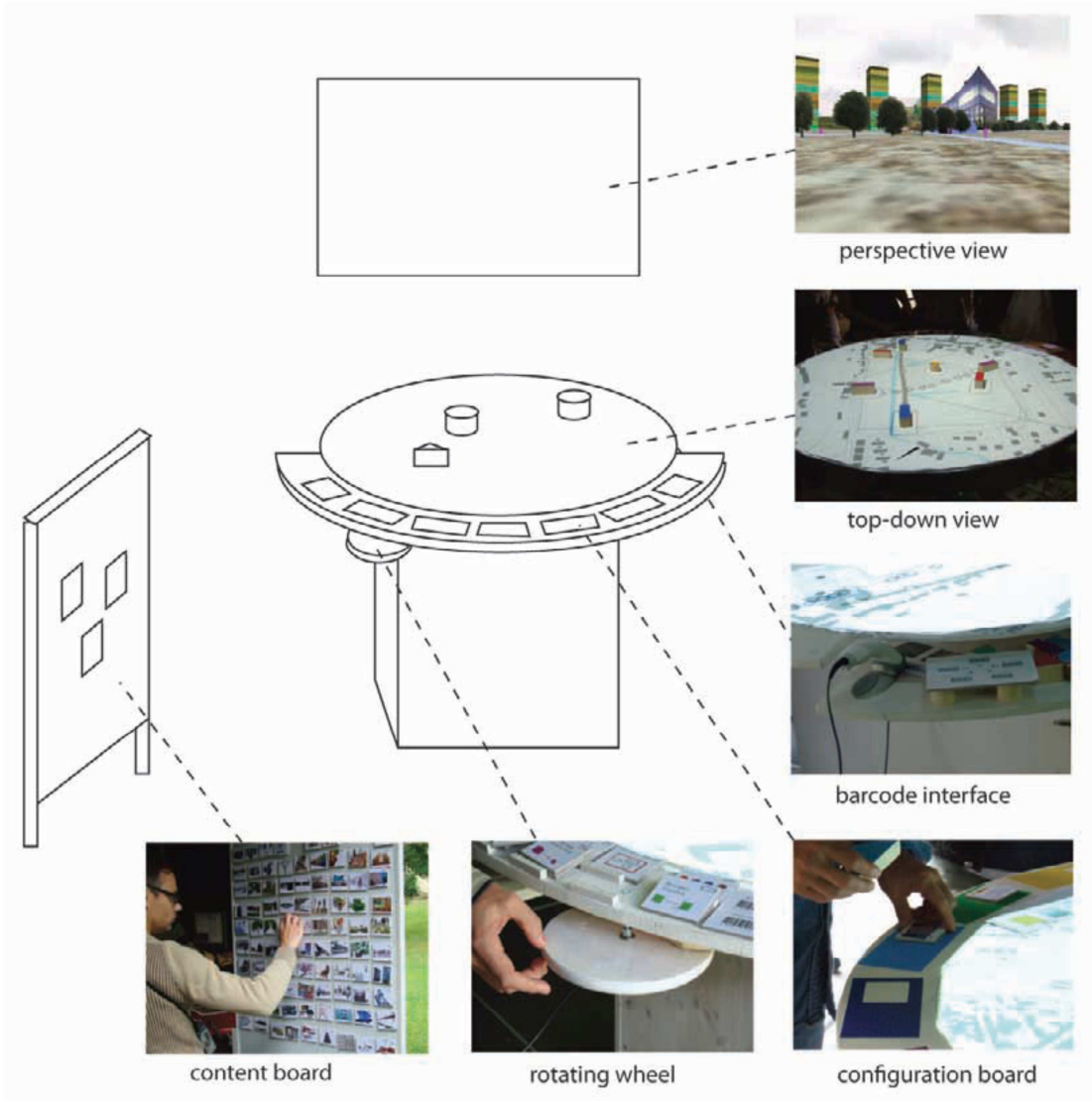


Figure 4: The components of the final version of the ColorTable [1]

2.2.2 The Tokens

The tokens are *"[…] physical handles to the digital content, and allow to position and manipulate it directly, by placing, moving and rotating them on the physical map"* ([5], p. 8) and they are reminiscent of the colored wooden building blocks that little children like to play with. Their outline is either that of a triangle, rectangle, or circle and they come in seven different colors. The color pink is reserved for the tokens that function as an eraser or are used to set a new viewpoint. The shape of the other tokens defines what type of content can be assigned to them. A camera above the table and an underlying color tracking system detect the color, shape, orientation, and position of the tokens.

Rectangular Tokens – Connections

The end points of a connection are two rectangular tokens that have the same color. The curve of the connection can be manipulated by rotating the tokens. The connections can either be streets or walking paths and the user can specify the traffic density and texture. The connections are animated with pedestrians and cars moving along them according to their type and traffic density.

Circular Tokens – Ground Textures

Ground textures, like grass, asphalt, stone, water, and so on, can be assigned to areas that are completely confined by connections. A circular token is placed anywhere in that area on the map and thus the texture associated with the token is allotted to the enclosed ground.

Triangular Tokens – 3D Objects and Billboards

The triangular tokens are used to place 3D objects and 2D billboards. The 3D objects are mostly simple geometric forms that can be combined and stand in for houses or other structures. Rotating the token also rotates the 3D objects. The billboards are 2D rectangular pictures of urban elements, like trees or flowers. They are always aligned in the direction of the viewer. A row of the same elements, both 3D objects and billboards,

can be created by placing two triangular tokens of the same color in the same manner as connections. As described above the attributes of the digital elements can be altered to fit the vision of the user.

2.2.3 Perspective Views

The ColorTable system offers different types of views onto the planning site and the interventions of the users. These visualizations are mixed realities.

The Panorama View

Before each workshop a number of cylindrical panoramas are created of the urban site from different viewpoints. The user can rotate or zoom into these photorealistic panoramic views and place virtual objects into them. The panoramas incorporate a depth map to realistically deal with occlusions and a height map that contains the height information for every coordinate in the area, thus creating 3D scenery.

The Video Augmented View

As the name suggests, the video augmented view is a live video stream of the real surroundings. The camera is mounted outside on a special apparatus and the users can pan and tilt the camera and zoom into the scene with an external controller. The virtual objects are perfectly aligned within this view.

Another camera can also be carried by a human scout in close proximity of the ColorTable. The users give verbal directions to the scout to guide him to areas of interest.

The See-Through View

An optical see-through screen is placed in front of a window that faces onto the planning site. The screen is a special kind of grid that reflects the projection, while the view behind it is still visible. The virtual objects are projected onto this screen augmenting the landscape beyond. Unfortunately, the aliment of the virtual objects in

the real environment can only be accurate from one specific viewing position and illumination contrast is also a problem.

2.2.4 Further Features

Urban Sketcher

The Urban Sketcher is Mixed Reality software that was developed at the Graz University of Technology [3]. The goal of the Urban Sketcher is to facilitate and improve communication in the context of urban design issues. The resulting software enables the user to:

- sketch and draw onto the panorama,
- insert and manipulate 2D and 3D geometric objects, and
- access and position further virtual content.

These functions can be selected and performed by either using a laser pointer on the projection screen or an API.

Besides the perspective views mentioned above the Urban Sketcher also provides access to an *aerial view* via a physical map that is captured with a webcam and a *token view* that can be navigated through with a special token.

Sound Application

In order to create a truly immersive virtual environment, objects and panoramas can have 3D sound connected to them. For example, pedestrians or vehicles that move along a flow have sounds associated with them which follow them around. The user can choose between three different hearing positions that are associated either with the camera view, a part of a flow, or can be randomly selected with a special *sound token*.

3 Gesture

3.1 Definition

Adam Kendon, one of the leading authorities in the field of gesture studies, defines gestures as follows: *"'Gesture'[…] is a name for visible action when it is used as an utterance or as a part of an utterance."* ([7], p. 1) In order to be able to employ this definition for the analysis of video material later on, a closer look at what constitutes an utterance is necessary, as well as which features discern visible actions as being relevant as part of an utterance.

Whenever human beings are surrounded by others, they unavoidably emit information through visible bodily actions. Consciously and unconsciously we share our intentions, state of mind, ideas, social status, and personal characteristics with others without using words. For example, a person's focus of attention and involvement can be easily deduced from the orientation of his/her body and direction of gaze. Moreover, there is no way to stop this flow of information, because even if one were to stand stock still in the presence of others, this behavior in itself would strongly invite interpretations about the intentions and state of mind of that person. [7]

Goffman [8] differentiates between information that people involuntarily *give-off* and information that people *give* through actions that are specifically designed to communicate. These actions or sequences of actions are treated as utterances. In other words, these actions are understood by the interlocutors as a meaningful, voluntary move, turn, or contribution in an exchange. Such utterances can consist of either speech or visible bodily actions or more often an interplay of both of them.

While speech appears to be instantly recognizable at all times, even if spoken in a foreign language one does not understand, the characteristics that determine whether we see a bodily action as an utterance are less straightforward. Pointing or waving in order to say good bye are always understood as gestures, but displays of emotion like crying or laughing are usually not as long as they are seen as genuine. Also, activities like playing with one's hair, biting ones fingernails, or adjusting one's clothing, even though one might deduce a person's state of mind from them, are not regarded as gestures. Furthermore, activities like smoking, drinking, or eating during a conversation are normally not understood by the interlocutors as having any informational intent even if they may somewhat regulate the timing of a conversation. [7]

But gestures can also come in disguises. Take for example the gesture known as *bras d'honneur*, a gesture that is regarded as extremely vulgar, so vulgar that performing it in public can land one in front of a judge. The nicest interpretation of it would be "Up yours!" Especially in France and Spain making that gesture is seen as highly offensive and deeply insulting and gets politicians [9] and soccer players [10] into huge controversies.

Figure 5: Bras d'Honneur [11]

The bras d'honneur is made by first straightening one arm with the inside of the arm and clenching the hand into fist. Next the forearm is brought up to form a right angle with the upper arm. The hand of the other than grabs the elbow of the bent arm. The entire gesture is performed in an upwards motion. [12]

Since the gesture is so offensive, people in Malta have altered the gesture in such a way so they can use it without being caught out. The disguised version is carried out by straightening the arm and the hand clenched into a fist. While the other hand lightly rubs the inside of the elbow of the outstretched hand. Thereby one seems to be merely taking care of an itch. [13]

The issue of gestures is further complicated by the fact that they can have different connotations in different cultural areas. The gesture where the thumb and the forefinger touch each other to form a circle and the rest of the fingers are extended straight away from the hand, as shown in Figure 6, is a good example for conflicting meanings. In America and in parts of Europe the gesture is completely inoffensive and is simply seen as a sign that everything is ok. But in South America and in regions of Southern Europe the gesture is understood as referring to a bodily orifice and therefore very far away from the meaning of things being ok. [14], [15]

Figure 6: OK Sign [14]

Even though the examples above show that gestures are not always straightforward, humans still intrinsically know which body movements are gestures. In 1978 Kendon [7] carried out a study to determine the features that a movement must have in order to be seen as a gesture. A film of a ritual performed by a tribe in Papua New Guinea was shown without sound to 20 English speaking participants from Australia. The participants were then asked to describe the movements they had seen and what significance they ascribed to the movements. All of the participants without exception first reported movements they saw as deliberate and part of the communication. Only afterwards did they describe movements that they considered natural or ordinary and

that did not carry any meaning. With only a few minor differences, the movements that were mentioned and the order in which they were reported adhered to the following categories:

1. Movements regarded as being deliberate and conscious attempts to communicate information.
2. Movements which were necessary to sustain or alter the body's position and orientation. This also included movements of the head and or eyes when seen as a switch of attention.
3. Movements that manipulated clothing or objects.
4. Movements that were not regarded as intentional and deliberate, but as unconscious expressions of nervousness or incidental. These movements were usually only mentioned after additionally prompting from the interviewer.

That the movements were almost unanimously reported in that order allows the conclusion that movements that are deliberately used to impart information receive the most attention compared to other movements. This finding is also mirrored in Goffman's concept of *attentional tracks*. [16] Goffman proposes that in any social encounter actions can be assigned to three different tracks regarding the kind of attention they receive:

1. *Main-line* or *storyline* track – activities that are central to the unfolding of the interaction and the main reason of the social encounter. This track only includes actions that "give" information (see above).
2. *Directional* track – *"a stream of signs which is itself excluded from the content of activity but which serves as a means of regulating it, bounding, articulating and qualifying its various components and phrases."* ([16], p. 210) For example, *batons* or *beats* are gestures that place emphasis on certain words or phrases in an utterance. [17], [18] They would be considered part of the directional track and will be described in more detail below.
3. *Disattend* track – activities that are typically human, for example scratching, rubbing ones eye, smoking, and which are accepted delineations from the

expected conduct in human interaction. This track may still carry information about the state of mind of an interlocutor.

Gestures are therefore movements that are part of the main-line track and which are always deliberate and communicative. Kendon [7], [19] was able to identify distinct characteristics of movement of gestures:

- The movements of a gesture are *excursions*, which means that they start by moving away from a position of rest and are always concluded by returning to the same rest position. There is no sustained change in body position.
- Gestures have clear boundaries. Their onsets and offsets are highly recognizable.
- Gestures follow a *peak structure*. The movement has a center that fulfills the communicative purpose of the gesture. This is also called the *stroke*.
- The movement is symmetric. Gestures look almost the same when played backwards as when they are played forward.

This list does not claim to be exhaustive and Kendon is certain that further research will find more kinetic features of gestures, but the given characteristics are distinctive enough to make a clear identification of gestures possible in the analysis of video material presented later in this thesis.

In summary it can now be concluded that gestures are deliberate and intentional visible bodily actions that communicate meaning. They are part of an utterance or an utterance in themselves and display distinct characteristics of movement. Thus we also have the tools to define gesture units for further examination.

3.2 Gesture Units

The description of gesture units that follows is based on [7] and [20]. Kendon provides only a short theoretic overview of the different phases of a gesture unit and his main

focus is on detailed examples, while McNeill gives detailed descriptions and only a short illustrative example.

The largest entity in Kendon's hierarchy is the *gesture unit*. The gesture unit corresponds to the interval between starting the movement with a certain body part from a rest position culminating in the peak or stroke of the structure and then returning to the rest position. Kendon identified three different phases during these movement excursions: *preparation*, *stroke*, and *recovery* or *retraction*. Later Kita [21] amended this scheme with *pre-* and *poststroke hold* phases. Furthermore Susan Duncan suggested to McNeill an additional *stroke hold* phase for the special case of motionless strokes.

Preparation Phase

In the preparation phase the limb starts in the rest position and is moved into the gesture space to the start point of the stroke. Once the limb is in place the movement comes to a temporary halt called the prestroke hold. The function of a prestroke hold is to delay the stroke until it can coincide with a specific spoken segment.

Stroke

The stroke can be considered the climax of the gesture unit and is the phase that carries meaning. The meaning of the stroke does not have to be identical with the co-occurring linguistic segment, but often provides additional complementary information. They express aspects of the same underlying idea in two distinctly different ways. The provided meaning can also overlap.

In a large sample of gestures, 90% of the strokes occurred simultaneously with the co-expressive speech. When the stroke and the semantically-linked speech segment were not in synchrony, the stroke usually takes place before the corresponding speech. The speech delay is mostly ascribed to short hesitations and the delay is small. In the rare cases were the stroke was preceded by speech, it was due to neurological anomalies.

A special form of strokes is a stroke hold. A stroke hold carries meaning, but there is no movement. For example a hand is placed into the upper regions of the gesture space in the preparation phase and held there to allude to the upper floor of a building.

If the movement freezes in the final position and posture of the stroke then this is referred to as the poststroke hold. It occurs when the motion of the stroke has finished, but the co-expressive speech segment is still in progress. Thus the poststroke hold, like the prestroke hold, fulfills the function of maintaining the synchronization of the stroke and the corresponding speech.

Recovery

The phase that concludes the gesture unit is the retraction or recovery phase. This consists of the movement of the hand back to the resting position. But the speaker can pass over the retraction phase and move directly into a new stroke.

Gesture Phrase

Kendon groups the preparation and the stroke with its corresponding pre- and poststroke holds into gesture phrases. A gesture phrase only contains one stroke, but a gesture unit can be made up of numerous gesture phrases. The recovery phase is considered a part of the gesture unit.

3.3 Link between Gesture and Speech

The examination of gesture units suggests that there is a strong link between gesture and speech. When gesture and speech occur together they express meaning together. They are co-expressive. Not only that, but at the points of co-expression they happen strictly simultaneously. This synchronicity of gesture and speech deserves special attention, because it leads to the conclusion that the brain is doing the same thing in two different ways at the same time and not two things separately. [20]

"Further observations show that such synchronized speech and gesture events comprise virtually unbreakable psycholinguistic units; unbreakable, as long as speech and gesture share meaning." ([20], p. 24) McNeill can substantiate his statement with the following observations that prove the strong bond of the gesture-speech unit. [20]

3.3.1 Gesture Synchrony and Delayed Auditory Feedback

During the *Delayed Auditory Feedback* (DAF) experiments the subjects continuously heard through headphones what they had said, while they were speaking, with a time delay of 0.25 seconds, which roughly corresponds to one syllable. This impeded their speech drastically. The speech became considerably slower. The speaker was hesitant and prone to drawling and metatheses[1]. Despite the resulting speech impediments the synchronicity between gesture and speech remained unaffected. Thus the bond between the two is strong enough to withstand the adverse forces of DAF.

3.3.2 Inoculation Against Stuttering

Mayberry and Jaques [22] made two significant discoveries regarding gesture production in connection with stuttering. When stuttering, involuntary repetitions of sounds, syllables, words, and phrases severely hinder the flow of speech. Unnatural prolongation of sounds and blocks that prevent the production of speech also characterize stuttering. [23] Mayberry and Jaques found that stuttering did not occur during the onset of a stroke. When the stroke was in progress and during the other gesture phases, stutters could still disrupt speech. If the stutter happened in a stroke, the expressive movement froze immediately and the hand either remained frozen in its position or returned to the rest position. The stroke continued after the stutter had passed. Both cases show that stuttering and the production of the part of the gesture that carries meaning are mutually exclusive. This is further proof of the strong connection between gesture and speech.

[1] Metatheses or Spoonerisms are specific errors in speech where consonants, vowels, and morphemes are switched. For example: "A blushing crow" – "A crushing blow" [46]

3.3.3 Gestures of the Blind

Drastic proof of a speech-gesture bond is the observation that congenitally blind people, who have never been able to see, gesture while they talk. Furthermore, Iverson and Goldin-Meadow [24] observed that the blind used gestures as often as seeing subjects. During their research they observed conversations of blind children and seeing children as well as blind children with other blind children. The children were always aware whether their interlocutors were blind or not, but having a sighted or blind interlocutor did also not affect the rate at which they gestured. This shows that thinking in gestures is not dependent on being able to see.

3.3.4 Information Exchange

Information that was conveyed by gesture may be later remembered as having been communicated by speech only. The subjects also did not remember that the information originally was presented to them through gestures. [25] Moreover, the same is also true the other way around. Kelly et al. [26] found that information delivered through speech was on some occasion recalled as information that was transmitted through gestures. This suggests that the bond between gesture and speech is symmetrical in nature.

3.3.5 Gestures and Fluency

A common opinion advocated by many scholars is that gestures replace speech. McNeill concedes that gestures can jump in when words fail, but this is not their main function. More often than not, rather than taking over when speech decreases, the occurrence of gestures also diminishes. Additionally, simple expressions in speech are accompanied by simple gestures and if the complexity in talk increases the complexity in gestures follows suit. When the fluency of speech decreases so do gestures. They start to rise again when fluency in speech is regained. *"The level of motor involvement via gesture parallels the fluency of speech and fullness of memory."* ([20], p. 27)

4 Categorizations of Nonverbal Behavior

The following chapter takes a broad look at different approaches to categorizing gestures and other nonverbal behavior. The work of David Efron was chosen, because his was the first attempt at studying gestures and it is a multifaceted approach to gestures. Paul Ekman and Wallace Friesen deal with gestures in the context of nonverbal behavior as a whole and from this it can be inferred what constitutes a gesture. Then David McNeill provides the framework to correctly assign them to categories.

4.1 First Cultural Analysis of Gestures

David Efron undertook the first comparative cultural analysis of gestures. Around 1940 he investigated the gestures of immigrants from Southern Italy and Yiddish speaking Eastern European Jews in Manhattan. The study included both first generation immigrants and descendants of both groups, who were already assimilated to a certain degree. The results of his examinations showed that gestural use is not biologically inherited but determined by culture.

While he did not strictly speaking define a typology of gestures as such, his detailed procedural method introduced new ways of looking at gestures and his work was indirectly very significant, because it influenced many classification schemes. Efron presented his investigation in [27]. The summary given below is based on centralization by Kendon in [7].

The focus of Efron's investigation was gestures preformed via hand and arm movements. He looked at them from three different perspectives, a *spatio-temporal*, an *inter-locutional*, and a *linguistic* perspective.

4.1.1 Spatio-Temporal Perspective

The spatio-temporal perspective looks at different attributes of gesture movements:

1. The extent of the radius of the movement and where on the arm the movement began.
2. Whether the form of the movement is sinuous, elliptical, angular, or straight.
3. Whether the plane of the movement is sideways or traverse and whether the gesture moves towards or away from the interlocutor.
4. The body parts that are involved in the gesture, for example, gestures that involve head movements, the positions of fingers during the gesture, was the gesture performed by one hand or both, and if two hands were involved, were they in unison or did the movement flow from one hand to the other.
5. Whether the progression of the movement was uniform or were there abrupt changes in the tempo of the movement. Efron referred to the later as *dischronic transitions* and the former as *flowing transitions*.

4.1.2 Inter-Locutional Perspective

The inter-locutional perspective is concerned with the involvement of the conversational partners. Did the interlocutors touch each other? Were there any instances were gestures occurred simultaneously? How did the participants position themselves and how much space was between them? It was also noted whether or not props were used to perform a gesture.

4.1.3 Linguistic Perspective

Within the linguistic perspective the meaning of gestures was analyzed. Efron distinguished between two different kinds of gestures.

Discursive Gestures

The first are gestures that are *logical* or *discursive* in their meaning. In other words, they do not impart any additional meaning independently from speech, but chart the ideational course of the conversation itself. They are either baton-like in nature establishing the beat of a conversation or weaving through the air outlining the direction and shape of the thoughts and ideas of the speaker.

Objective Gestures

The other type of gestures Efron examined was called *objective* gestures. These are gestures that express meaning in a self-contained manner independently from speech. They can occur simultaneously with speech or completely on their own. Efron breaks this category further down into *deictic* or pointing gestures and *physiographic* gestures. Physiographic gestures refer to objects or ideas and their shape takes the form of depictions or characterizations of their referent. Efron further distinguishes between three different types of gestures in this category: *iconographic*, *kinetographic*, and *symbolic*.

Iconographic gestures are gestures that demarcate the shape of an object, explain spatial relationships, or indicate the relative size of something. In contrast, kinetographic gestures mimic different bodily actions, like performing exaggerated typing movements with ones digits to emphasize that one has been busy typing or using ones index and middle finger to imitate a walking movement.

With symbolic or *emblematic* gestures the gesture is either pictural or nonpictural and it does not have any morphological tie to the idea or object that it expresses. Their meaning tends to be standardized within a specific cultural tradition. An example given by Efron for this type of gesture is the Jewish gesture for “impossibility”. This gesture is

performed by poking the flat palm of one hand with the index finger of the other hand, implying that something will happen only if grass starts to grow on the palm of the speaker's hand.

4.2 Categorization of Nonverbal Behavior

In their article "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding" [17] Ekman and Friesen laid out a road map to fully understand a person's nonverbal behavior. At the time of publication the study of nonverbal communication and nonverbal behavior was in its infancy and the paper was a great success within the scientific community. [7] While strictly speaking the term "gestures" is not used in the article, two of the categories described contain gestures as they fit the definition given above. The categories are called *emblems* and *illustrators*, which have been very extensively referenced.

The article is the culmination of different research projects conducted by Ekman and Friesen. Among them are studies concerning how observers deduce a person's attitude, personality, and mood from facial expressions (mainly in the setting of a psychiatric facility), cross-cultural comparisons of facial expressions and gestures, and how deception can be detected from nonverbal behavior. In their research they use a very broad and all-encompassing definition of nonverbal behavior that includes "*[..] any movement or position of the face and/or the body […].*" ([17], p. 49)

This wider look at nonverbal behavior is essential to comprehending the difference between gestures and nonverbal behavior that does not explicitly impart any meaning. Having a well-founded understanding of the distinctions between the two is a valuable tool for the analysis of the video material.

4.2.1 Five Categories of Nonverbal Behavior

Based on the findings of Efron [27] and their own research, they deduce five categories of nonverbal behavior:

1. Emblems
2. Illustrators
3. Affect Displays
4. Regulators
5. Adaptors

Emblems

Emblems are a specific subset of gestures. They are nonverbal acts whose meaning can be directly translated into one or two words or a phrase. The definition of an emblem is known by each person within a certain social and cultural group. An emblem may repeat, replace, or contradict the verbal information it occurs with simultaneously. An interlocutor is usually aware that he/she used one and thus they are an intentional and deliberate act of communication. Therefore emblems are to some degree chosen to fit the external circumstances. On rare occasions, especially in heightened emotional states, people can be unaware that they displayed an emblem. This is comparable to a slip of the tongue.

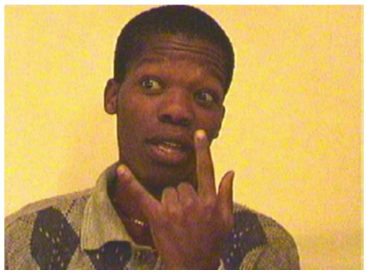


Figure 7: Clever Gesture [28]

An example for an emblem is the *Clever Gesture* described by Brookes in [28]. The gesture is widely used by black male youths in South African townships. The general

form of the gesture is the ring and middle finger folded in and the remaining fingers extended outward (see figure above). To perform the gesture the palm faces inward. The tips of the index and little fingers are directed towards the eyes and the hand is moved in a side-to-side motion in front of the face. Depending on the situation and slight variations in usage the gesture has a number of different connotations. When it is used in conjunction with language the gesture means that someone or oneself is "street-smart". Being able to quickly take a situation and to react accordingly is an invaluable skill in a township. When the gesture is used independently and with slight changes in movement it refers to a range of meanings that have at their semantic core the concept of "seeing", for example, *"[…] 'I want to see you,' 'Look,' 'Be alert,' 'Watch out,' 'You are being watched,' or 'I see you' as a greeting."* ([28], p. 19)

Additionally, variations of this gesture also have very different meanings for members of other groups. In American Sign Language the gesture turned outward is an acronym for "I love you". [29] If the thumb is bent inwards and rests on the index and middle fingers the gesture is the sign of the horns also called *corna*. Its usage ranges from warding off evil to implying that the recipient's wife is cuckolding him depending on the cultural surroundings. Furthermore the sign was very popular within heavy metal subculture with a much darker satanic connotation, until it went mainstream in America and was watered down to simply mean "rock on". [30], [31]

Illustrators

Illustrators are the second and last category dealing directly with gestures. This type of nonverbal behavior is closely linked to speech. Their purpose is to further illustrate verbal information. Based on Efron's work, Ekman and Friesen distinguish between six different types of illustrators, five of which have been discussed above: *batons, ideographs*, *deictic movements*, *spatial movements*, *kinetographs*, and an additional new type, *pictographs.* The referent of a pictograph is drawn like a picture into the air with the movement of a body part.

The relationship between illustrators and the simultaneous occurring speech is incredibly tight and content, infliction, and loudness of what is said can influence the

usage and execution of illustrators. The information presented by them can duplicate, replace, contradict, or supplement the spoken meaning. In comparison to emblems people are slightly less conscious about their usage of illustrators. Efron also found that the types of illustrators used depend on the social and cultural background of a person. [27]

Affect Displays

Another type of nonverbal behavior displays emotions. Feelings are primarily shown on the face. Body movements like trembling or shaking can occur with certain emotions, but are usually more a way of dealing with the affect than showing the affect itself. This category does not include gestures.

Many researchers agree that there are *primary affects* that are the same for all members of the human species. In other words, certain movements of facial muscles are read by everyone as the same emotional state. Ekman and Friesen provide a tentative list of seven primary affects: happiness, surprise, fear, sadness, anger, disgust, and interest. Their research has shown that each of these affects can be identified by any person regardless of their cultural background.

The muscle structure of the face is complex enough that different feelings can be displayed simultaneously and emotions are often not displayed solely on their own. Affect displays of numerous emotions can happen simultaneously or in quick succession. Both cases are called *affect blends.* The structure of an affect blend is governed by the evoking circumstances or acquired habits of correlating one affect with another.

Regulators

As the name suggests this category includes nonverbal behavior that regulates turn taking in a conversation between two or more interlocutors. Regulators can either let the speaker know that he/she should adjust the speed of his/her speech, repeat something, provide more information, be less salacious, let somebody else speak, etc. or be directed at the listener to make him/her pay more attention, wait a little longer for

his/her turn to speak or indicate that he/she can talk now. Regulators are usually minute nonverbal acts, for example, eye contacts, slight changes in body posture, lifting of eyebrows, or changing the distance between interactants. The head nod is the most common regulator.

The function of the regulators is to structure the flow of a conversation; therefore, they do not convey any meaning that is relevant to the content of the conversation as such. The difference between batons and regulators is that batons add emphasis to a word or chart the flow of an idea within a conversational turn, while regulators structure the entire exchange between interlocutors. Any act of nonverbal behavior can have a regulatory function, but only actions whose only purpose is to regulate a conversation are considered to be regulators.

Informal research conducted by Ekman and Friesen has shown that people do not consciously use regulators to structure the conversational flow and similarly the other interactants are not aware of reacting to them. Most people also find it very hard to omit them, and if they are absent the interlocutors become irritated and often the conversation completely breaks down.

Adaptors

Adaptors are the last category of nonverbal behavior and also a rather diffuse category. Mostly in childhood, humans learn adaptive patterns that fulfill very basic human needs, such as methods to deal with emotional or physical discomforts, to establish and maintain interpersonal contact, or learn fundamental tasks. When these adaptive efforts are first established in a human, they are associated with specific urges, emotions, expectations, types of interpersonal interaction, or a certain situation.

Adapters are residue of these behavioral patterns that over time have become fragmented. When they appear in adult conversation something related to the original association triggers them, but without knowing the trigger and because not the full adaptive pattern is carried out, the behavior might only be perceived as random or noisy body movements. The person displaying adaptors is usually unaware of them,

because they are habitual to her/him and thus they are not intended to be communicative.

Adapters can be broken down into three different categories depending on the object of the adaptive behavior. An adaptor can either be directed towards the own body of the individual, somebody else, or an object. Accordingly they are called *self-adaptors*, *alter-directed adaptors*, or *object-adaptors*. Behaviors that maintain or reestablish the bodily well-being of an individual are self-adaptors. They can stem from a variety of sources, like behavior that was learned to fulfill bodily functions, block or enhance sensory inputs, or upholding personal appearance to meet social standards. They are all usually learned in childhood and shaped by social conventions. An example would be the repeated licking of the lips even though the lips are not dry, but when the person is satisfied with something similar to enjoying a good meal, but there is also no food present. Grooming in public would be another example, but often happens with both conversational partners being aware of it, but ignoring it out of politeness. This is the case for most self-adaptors. People rarely comment on them, but rather pretend that they have not seen them.

The alter-directed adapters stem from movements learned in early childhood that physically relate the body to other bodies. This includes the passing of an object between two people, defending oneself from attack or attacking somebody, or building and perpetuating affection and physical intimacy. Distinguishing alter-directed adapters from other nonverbal behavior can be very difficult. For example, crossing the hands in front of the body can either be to ward of the cold and hence a self-adaptor or to protect the body from a real or imagined attack and then would be regarded as an alter-directed adapter.

An object-adaptor is behavioral fragment that was part of a course of action that was learned in order to perform a mechanical task, like smoking, steering a vehicle, wielding a hammer, knitting, and so on. This type of adaptors differs from the others in that they are acquired later in life. The speaker is often aware of performing them and if that is the case they can be also used to actively communicate.

4.3 Categorization of Gestures

Another classification scheme was developed by David McNeill and Elena Levy. The scheme was first proposed in [32], later elaborated on in [18] and summarized in [20]. The focus of McNeill's research is explicitly on idiosyncratic gesturing that occurs spontaneously with speech, because he believes that gestures and speech are one integrated system and reveal information about the thought processes of a person. McNeill also takes the viewpoint that gestures need to be analyzed regarding their function to illustrate the content of speech.

During his research, data was gathered by showing participants a short Sylvester and Tweety Bird cartoon or the feature film *Blackmail* by Alfred Hitchcock. After viewing the cartoon or film the test subjects were asked to narrate the film to another person or answer a set of questions about it. These conversations were videotaped for later analysis. The participants, ranging from children to adults, were told that the purpose of the experiment was to research storytelling so to avoid them becoming conscious of their gestures. [18]

4.3.1 Category System

Imagistic and Non-Imagistic Gestures

Based on his research McNeill's first level of distinction is between *imagistic* and *non-imagistic* gestures. Imagistic gestures convey a picture of some sort. These types of hand movements might for example outline the shape of an object or represent a certain pattern of actions. Non-imagistic gestures on the other hand are for example pointing gestures or gestures that mark the rhythm of a conversation, like the batons defined by Ekman & Friesen.

Iconic Gestures

Imagistic gestures can be further differentiated between *iconic* and *metaphoric.* McNeill defines iconic gestures as "*[…] gestures in which the form of the gesture and/or its manner*

of execution embodies picturable aspects of semantic content (aspects of which are also present in speech).” ([20], p. 39) In other words, the hand movement draws an image of a specific action or object into the air. As an example McNeill describes a gesture that illustrated the description of bending back a tree by a participant of his research (see figure below).

Figure 8: Iconic gesture - And he [bends it way back] [18]

The participant grips something in the air above his head and then bends his arm towards his shoulder following an arch trajectory. The execution of the gesture took place simultaneously as the participant uttered the phrase *“bends it way back”*. The gesture realizes its symbolic function by resembling the action described. The arch of the hand movement simulates the action of bending a tree. The information provided by the gesture and the speech is not redundant even though it often overlaps. Since the gesture was carried out with one hand it is suggested that the tree was fixed into the ground on one end, a piece of information not given in speech. Gesture and speech both work in unison to express a concept. They are *co-expressive.*

Metaphoric Gestures

While gestures can depict tangible actions and objects they can also impart *images of the abstract*. These metaphoric gestures have an iconic component in that they are pictorial. McNeill uses the following gesture to explain this concept (see figure below).

Figure 9: Metaphoric Gesture - It was a Sylvester and Tweety cartoon [18]

The gesture occurred at the beginning of the description of the cartoon and the participant is explaining to the listener, that he will be describing a cartoon. He has the hands raised up in parallel and the palms are facing each other as if he were presenting an object to the listener. In a way he does just that, the gesture has turned the abstract idea of the film genre cartoon into the concrete form of a bound object that he displays and offers to the listener. The iconic gesture of holding and presenting something becomes a metaphor for presenting an abstract idea or concept. Furthermore, the gesture space itself can be used to create a metaphor. In one instant a participant divided the space in front of him into a spectrum of morality to delineate the moral quality of the story characters.

Beats

The next category of gestures is what McNeill calls *beats*. They are basically the same as the batons described by Efron as well as Ekman and Friesen in that they emphasize words or phrases and outline the flow of ideas, but McNeil places more significance on the second aspect. *"Beats reveal the speaker's conception of the narrative discourse as a whole. The semiotic value of a beat lies in the fact that it indexes the word or phrase it accompanies as being significant, not for its own semantic content, but for its discourse pragmatic content."* ([18], p.15) He found that people use beats amongst other things to mark the first mention of a new character, highlight a change in topic, or summarizing

what happened so far. Thus beats accompany information that do not further the storyline, but rather constitute the structure within which the storyline can take its course.

Deictic Gestures

Deictic or pointing gestures round off what McNeill has dubbed the Iconic-Metaphoric-Deictic-Beat Quartet in [20]. Pointing is discussed in more detail later on and McNeills findings will be presented in that context.

Cohesives and Butterworths

McNeill excluded two categories that he described in [18], namely *Cohesives* and *Butterworths.* Cohesive gestures link together parts of the discourse that thematically belong together, but are temporally separated. The cohesion is established by repeating the same gesture. Butterworths, named after the British linguist Brian Butterworth, are gestures that the speaker uses to retrieve a word or phrase from memory. Both types of gestures use the other gesture types for their purposes. This might be the reason McNeill did not include them in the later gesture type summary, because they only add a function to the other four types.

Kendon's Continuum

McNeill found during his research that most gestures are too complex to fit in just one category. Gestures are a composition of the different features and should be analyzed according to what degrees they display *iconicity, metaphoricity, deixis, "temporal highlighting"*, and so on.

To provide a tool for this kind of analysis, McNeill developed "Kendon's Continuum". In [33] Adam Kendon categorized gestures according to the part they played in an utterance. McNeill placed these categories along a continuum and named the result after Kendon to honor his groundwork [18]. After some consideration McNeill split the first continuum into four continua in order to capture the full complexity of gestures [20].

4.4 Summary

Efron's meticulous examination of gestures is the first of its kind and teaches us that in order to fully comprehend gestures we need to look at them from a number of different perspectives. Therefore gesture classifications are numerous and depend on the perspective from which they view gestures. Especially his take on gestures from the linguistic perspective offers great insights into how gestures and speech co-express meaning. The significance of his work is also felt in the works of Ekman and Friesen.

Since Ekman and Friesen's research encompassed nonverbal behavior in its entirety, they do not define useful categories for the quantitative analysis. For example, the categories of object-adaptors and kinetographic illustrators strongly overlap and Ekman and Friesen do not provide a clear delineation between them. To make matters even more complicated, nonverbal behavior can jump from one category to the next. Affect displays can become emblems. In turn, illustrators can take over emblems to illustrate. [7] But they provide criteria to distinguish gestures from other nonverbal behavior that is not relevant for the analysis.

Even though McNeill believes that gestures should be measured with continua, using any of the four continua he proposes for analysis requires an enormously detailed and time-consuming examination of gestures. Such an analysis would greatly exceed the dimensions of this thesis because of the large quantity of material to analyze. A finer analysis of short sequences would be possible with more specificity.

For the scope and objectives of the quantitative analysis at hand placing gestures into categories is both sufficient and an efficient way to proceed. The categories of iconic gestures, beats, and deictic gestures describe very clear and recognizable features that can be easily applied to the gestures found in the video material and categorization can be performed even by a non-linguist.

We now have the tools to classify gestures for the analysis. The work of Ekman and Friesen provides the basis to identify gestures and McNeil describes the visual criteria

to correctly categorize them. The category of deictic gestures will be further broken down into special forms of pointing.

5 Pointing

Pointing or deictic gestures are such an integral aspect of collaboration in the context of the ColorTable that a closer examination of pointing is necessary. The most common form of pointing is done with the hands, but any other body appendage, like an elbow, a foot, the head, nose and eyes can be used. In some parts of the world even the lips are used to indicate something. Besides body parts, held objects can obviously also be utilized for pointing. [20]

In North America the prototypical pointing gesture is the G hand shape ☞. The index finger is outstretched and the remaining fingers are curled inward towards the palm. [20] The target of the point is identified by drawing an imaginary straight line from the furthest point of the extended appendage outward into the surrounding space. The deictic gesture can be directed towards a real object or location in the visible surroundings or somewhere outside the visible field. A speaker might point to a real landmark through a wall. In general the target can be an object, a location, or a direction, both real and imagined. [7]

Depending on the context the interpretation of the deictic gesture might be impossible without the accompanying speech, but pointing can function on its own as well. [20] But as the term deictic gesture suggests, pointing has a close relationship with deixis. Deixis are linguistic features or expressions that tie an utterance to its spatial and temporal circumstances. In many situations this pivotal bond cannot be established without some non-linguistic action and in a lot of those situations the only way to establish this bond is with pointing. [7]

5.1 Form of Pointing

Kendon collected a few studies, for example [34] and [35], that suggest the form of a pointing gesture also has an influence on its meaning. Pointing at a person with the index finger has the connotation of being authoritative or accusing. Using the head or thumb to indicate somebody can be regarded as impolite. If somebody would want to identify a person to another in secret an inconspicuous way is to use a gaze and a slight movement of the head.

Not only which appendage is used for pointing is relevant, but the shape of the hand as well. In [7] Kendon presents a comparative study of different hand shapes that he and his colleagues observed during the analysis of their video material of British and Italian speakers. They concluded "*that the form of pointing adopted by a speaker is systematically related to the way the object being referred to is presented in the speaker's discourse*." ([7], p. 201) That means if a speaker points to an object because she/he wants to illustrate or exemplify a point the hand shape is different from the one she/he would use to identify a specific object. The form of pointing indicates how the speaker wants the object in question to be considered. Kendon differentiates seven different hand shapes regarding the position of the fingers and the alignment of the forearm:

1. Index Finger Extended Prone (palm down)
2. Index Finger Extended Neutral (palm vertical)
3. Thumb Extended (orientation of forearm is variable)
4. Open Hand (i.e. all digits extended and adducted) Neutral (palm vertical)
5. Open Hand Supine (palm up)
6. Open Hand Oblique (forearm supination partial, palm of the hand faces obliquely upwards)
7. Open Hand Prone (palm of the hand faces downward or away from speaker, depending upon flexion of the elbow or extension of the wrist)

Index Finger Extended

The *Index Finger Extended Prone* type is most often used to identify a specific object, place or person for the first time and is typically accompanied by an explicitly deictic word like *"this", "that", "here"*, or *"there"*. After the point of reference is successfully established the shape of any additional pointing changes to *Index Finger Extended Neutral*. This is the case if additional information is provided about the object or the object triggers or influences an event. Furthermore if another object is distinguished that is in some way related to the first object, this is done with the Index Finger Extended Neutral. These ways of using the Index Finger Extended are very common, but deviations have been observed.

Open Hand

In comparison there is a strong consistency in usage of the Open Hand for pointing. Whenever the Open Hand comes into play whatever is being indicated is not the primal focus of the conversation, but is in some way related to the main subject matter. For example, it might be an example for a certain type of objects, be the staging area for an activity that is being discussed, or set the stage for the main topic. In contrast to the Index Finger Extended the Open Hand is also significantly less often used in connection with a deictic word.

The *Open Hand Supine* belongs to a family of gestures that are characterized by embodying the semantic themes of either "presenting" or "receiving". The speaker might present something that warrants further attention or should be looked at. He/she might also point at an interlocutor in agreement with what the interlocutor said, in other words "receiving" the opinion of somebody else.

If something is pointed out with the *Open Hand Oblique* this is most commonly done when a comment is being made about it, either directly about the object or how it pertains to the speaker. In most cases the remark is directed towards a person and is negative. Typically, the critical remark is made in the presence of a 3rd party and is not spoken to the subject of the comment, but the 3rd party. The Open Hand Oblique directs the focus to the target of the criticism.

The last of the Open Hand pointing gestures is the *Open Hand Prone*, which takes the following form. The forearm is in a prone position and the wrist is extended in a way that the palm is turned away from the speaker and oriented vertically. This form was observed when the interlocutor talked about the spatial scope of an object being referred to or when a cluster of objects were being regarded as an ensemble.

Thumb Extended

If the thumb is used to indicate an object, the objects are either situated behind or to the side of the speaker. This is partly due to anatomical presuppositions in that the thumb is naturally curved backwards when extended. It would be rather awkward to compensate this, when using the thumb to point forward. The thumb is not exclusively used to indicate something behind the speaker, the index finger is also used, but the circumstances vary. The index finger is used in the cases described above, but if the exact position of an object is unimportant or well known by all involved the thumb is used. This is also the case if the focus of the discourse has shifted away from the indicated object. Pointing with the thumb happens mostly parenthetically and casual.

5.2 Conditions for Meaningful Pointing

As an ethnographer, rather than a psychologist or a linguist, Charles Goodwin takes a much broader approach to gestures and especially pointing. He looks at them in the larger context of the activity in which they are embedded and in order for the pointing gesture to be a fully successful communicative act certain preconditions have to be in place. Besides the basic requirement that two or more people are in attendance, Goodwin [36] delineates a minimum of five semiotic resources that combine to produce a meaningful point:

1) *a body visibly performing an act of pointing*
2) *talk which both elaborates and is elaborated by the act of pointing*
3) *the properties of the space that is the target of the point*

4) *the orientation of relevant participants toward both each other and the space that is the locus of the point; and*
5) *the larger activity within which the act of pointing is embedded.*

[36] p.2

What sets Goodwin's approach to pointing gestures apart from those of linguists like Kendon is that he understands them as a *situated interactive activity*. In other words, he takes into account the activities and events that surround a pointing gesture and the semantical influences these activities have on the meaning of the surrounding space. This methodology connects the qualitative analysis and quantitative analysis that will be presented later, as it puts the pointing gestures in relation to the activity categories. Hence, conclusions can be derived from the usage of pointing gestures in the activity categories.

Kendon [7] calls the body from which a pointing gesture originates an *origo*, a term coined by Bühler. [37] The origo is the reference point from which vis-à-vis in space the referent of the deictic gesture can be located. The space where the target of the point can be found is called *domain of scrutiny.* If the interlocutor is oriented in such a way that she/he is unable to see the domain of scrutiny the speaker will pause both the conversation and the gesture until the addressee has repositioned herself/himself. The speaker has to remain constantly aware of the conduct of the addressee and might also have to adapt his/her position to allow for an unobstructed view. [36]

In their study of workplace deixis Hindmarsch and Heath [38] found numerous examples of this kind of interplay between colleagues in a telecommunications control center. In one instance a right-handed person used his left hand to point and shifted his body towards his interlocutor to make sure he had a clear view of the document he was referring to. Other examples included adjusting the alignment of a computer screen, rearranging documents, or waiting until a colleague arrived at the screen before finishing a pointing gesture and continuing to speak. The direction of the gaze of an interlocutor is often a very strong indicator for whether the same domain of scrutiny was successfully established.

The accompanying talk also facilitates the establishment of a mutual referent. Deictic terms, like “this”, can contain crucial information for identifying the referent. Distinctions can be made if one or more objects are being indicated (“this” vs. “these”) and in the case of “here” or “there” the focus of attention might not even be an object, but a location and the deictic terms also include whether it is close by or farther away. [38] [36] The verbal references can also be more specific by verbally referencing directly what particular type of entity is being pointed at, like a pen or stain on the carpet. [36]

But localizing the target of the point might not be enough to fully comprehend a communicative act. The domain of scrutiny can have its very own semantic connotations. Take for example a map, where the squiggly lines and different marks have assigned meanings. Pointing at a line on a map is mostly not done to point out a printed line on a piece of paper as such, but to indicate some geographical feature, like a river, a road, or a border. Furthermore, one needs to know the legend to the map in order to make sense of the geographical markers. [36]

But the semantic structure of a place may change with the encompassing activity that is carried out there. A part of Goodwin’s research focuses on the collaboration of archeologists on a historical excavation site; in this setting, different colored patches of dirt have very specific meanings for the archeologists. The remains of a decaying wooden post might just be a tube of dirt whose only defining feature is the difference in color to the surrounding soil. These kinds of phenomenon would have no meaning to a layman, but are the main focus of an archeologist’s work. The archeological term for this is *feature*. [36] [39]

Goodwin calls gestures whose meaning depends on their surroundings, where the term “surroundings” includes both space and activity, *environmentally coupled gestures*. It should be noted that other types of gestures besides pointing can be environmentally coupled. Goodwin gives an example of a participant describing the missing parts of a blender while holding a central piece of said blender. The description would make no sense with just speech and gestures alone. [39]

At first glance, it might seem as if pointing gestures are automatically environmentally coupled, but this is not the case. McNeill [20] observed that almost all of the pointing in adult conversations and narrations are not directed at real objects or places in the environment but is instead *abstract pointing*.[2] "*Abstract pointing is a species of metaphoric gesture, where space and a locus within it are used to present a nonspatial meaning.*" ([20], p. 40) In comparison to concrete pointing, which locates references in space, abstract pointing constructs new references in space. In a child's development concrete pointing is normally mastered before the child's first birthday. The emergence of abstract pointing does on average not happen until the age of twelve and marks an important stepping stone in the child's development.

The example given above for metaphoric gestures where the speaker turned the gesture space into a classification scheme for morality was achieved with abstract pointing. The central space was the reference point for perceived morality and the left space for actual morality. The speaker explained that one character at first seemed to be on the good side and pointed to the central space, but then she turned out to be the killer. This revelation was accompanied by pointing to the left space. Thus the gesture space was momentarily converted into a metaphor for the appearance/reality contrast.

5.3 Special Cases of Pointing

5.3.1 Dual Point

Goodwin [36] observed an interesting type of pointing during his research at an archeological excavation site. The gesture data was collected during the first couple of days after the excavation site had been closed for a period of time. In order to protect the site from environmental influences exposed surfaces were covered with a layer of

[2] This finding should be considered with certain reservations, because McNeill's research does not include conversations that take place around actual activities.

dirt when the last digging season ended. Thus when reopening the site, the first steps were to carefully remove the protective dirt and then to identify archeological features, which have been meticulously documented with maps, photographs, and so on.

In the relevant example, one of the graduate students working on the site believes that he has successfully identified a feature in the dirt. So he attracts the attention of a senior archaeologist to confirm his findings. He first points out the feature on a map with a trowel and then locates the feature in the soil with his gaze. The pointing is accompanied by the spoken phrase: "I think I finally found this feature." The deictic term "this" has two referents in the surrounding environment. The target of the point is present in two different places, once on the map and once in the dirt. The feature was first pointed out with the trowel on the map and then with the gaze in the soil. With this *dual point* both locations are incorporated. As a result, a semantic link is created between the two. The senior archeologist had to pay attention to both spaces designated by this complex point in order to be able to judge the situation correctly.

5.3.2 Collaborative Pointing

It is crucial to any pointing act that the addressee locates the intended target of the point. Body orientation and direction of gaze of the addressee are strong indicators for the speaker as to whether or not the referent has been correctly identified. Another way to demonstrate understanding is through c*ollaborative pointing.* The addressee demonstrates her/his understanding by also pointing at the referent at approximately the same time as the speaker. [36]

During his observations of the archeologists at the excavation site Goodwin found that collaborative pointing is often accompanied by the addressee providing verbal affirmation of having understood what the other was pointing out. In some cases this affirmation even overlapped with the talk of the other and had essentially the same meaning. [36]

Furthermore collaborative pointing was also a way for one of the students to demonstrate that she had understood the instructions of a senior archeologist and thus mastered the skill of identifying archeological features. The student anticipated where the senior archeologist would be pointing and outlined a feature with her teacher. "*Collaborative pointing thus constitutes an elegant solution to the problem of how to mutually demonstrate that each participant can independently parse a complex visual field into the specific phenomena that are relevant to the accomplishment of the task at hand.*" ([36], p. 27)

5.3.3 Tracing

Tracing blurs the boundaries between iconic and deictic gestures, because it is a mixture of both. When identifying features, as in the example above, Goodwin's archeologists would often trace their shape with a moving point thereby superimposing an iconic representation on the deictic nature of pointing. Tracing is done by shifting the pointing gesture along the outline of an entity and thereby highlighting its shape. Thus the pointing gesture mirrors the shape which gives the gesture iconic properties. [36]

Architects are another profession that regularly employs tracing. In [40] Murphy analyzes the gestures during an architectural design meeting and focuses on an example of tracing. During the meeting one of the participants traces a ramp on a floor plan. The deictic component of this gesture is that it indicates the ramp on the plan. By tracing the ramp, the length and shape of the ramp is displayed, which adds an iconic component to the gesture.

Furthermore, Murphy describes an incidence of what could be called *abstract tracing*[3]. Instead of outlining a concrete entity on the plan, one of the participants used a tracing gesture to illustrate an imagined path people would follow when delivering materials into the structure. The gesture represents an imagined activity in a building that only

[3] Neither Murphy nor Goodwin use this terminology in regards to tracing.

exists on paper, but it has to adhere to the constraints imposed by the walls, floors, and doors that are shown on the plan.

To sum up tracing, Goodwin [36] describes three main characteristics of tracing:

1. A dynamic relationship exists between the moving finger and the shape that is traced, in which both are needed to make sense of the other. Pointing does not stop after the referent as such has been established, but continues on.
2. Tracing provides complex information about the shape and size of the referent that would be very difficult to express verbally.
3. Tracing gestures have a strong link to their physical surroundings.

6 The Workshop

The IPCity technologies were tested and presented to the public in a series of workshops. In June 2009 a participatory workshop was held in the city of Pontoise, France, in collaboration with the City of Pontoise, the Agglomeration Community of Cergy-Pontoise, and the University of Cergy-Pontoise. The urban project at the center of this workshop was the future use of the Chamber of Commerce of Versailles (CCI).

The Chamber of Commerce of Versailles will soon relocate and the CCI building will become vacant. The building is situated close to the city center of Pontoise and sits on a major axis between the city and the public garden of Lavandières. Thus, the site has the potential to become an important connective link. The area of the CCI building also includes a parking lot and a garden. Another urban element that has to be taken into account is the Marcouville Estate. The park and chateau de Marcouville are situated to the West of the garden of Lavandières. The local authorities are planning to reopen the park (it has been closed to the public for security reasons since WW II). The future purpose of the chateau is also under discussion. Figure 10 is a map of the CCI building's location and the surrounding areas.

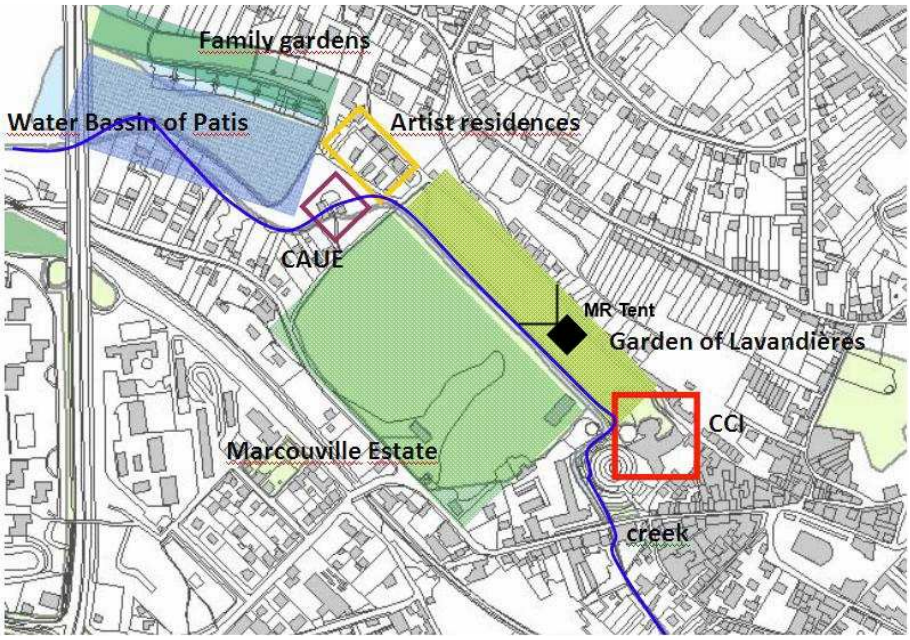


Figure 10: The location of the CCI building and the surrounding area [5]

From a broader perspective, the planning site belongs to the centralities that constitute the city of Pontoise. These centralities include, for example, the historic part of the city, the river Oise, the center of the new town of Cergy-Pontoise, the social housing neighborhood of Marcouville, the military barracks of Caserne Bossut and the St. Martin exhibition center. The workshop also dealt with how the project site can influence the connectivity within this network and valorize the other centralities.

Approximately a month before the participatory workshop, a preparatory urban workshop was held to give the participants the opportunity to get to know each other and to develop initial ideas. In this setting the ColorTable did not play a part, but traditional methods and representational tools were used to stimulate the imagination of the participants. The results guided the researchers in the preparation of content material for the participatory workshop and the workshop also serves as a comparative example to the work done with the ColorTable.

The participatory workshop lasted two days and took place directly at the site in the MR-Tent. Each day was a separate workshop with a different group of participants. Both groups included stakeholders of the site, urban planners, designers, city officials,

and residents. Several of the participants had experience working with the ColorTable from previous workshops and they could assist the others with how to use the different functions and tools of the ColorTable.

This analysis focuses on the 2nd day of the workshop and the following people participated on that day:

- **E:** A young urban planner, who is not directly involved with the official planning process for the site; he provides expert knowledge regarding urban issues and has worked with the ColorTable before.
- **EV**: Works at the CCI building and knows the site very well; because of her field of work she is familiar with commercial dynamics of the site and the animation of public spaces in general.
- **G**: A citizen of Pontoise; he is familiar with the site because he has a plot at the Patis family gardens.
- **Ch**: Head of the green spaces services of the City of Pontoise
- **S**: An artist who lives at the artist residences near the site. He brings his artistic sensibility and has experience with everyday life at the site
- **J**: A resident and a representative for the local association
- **M**: Moderator of the workshop

Interventions of the technical support staff and the researchers during the workshop were also included in the analysis. They are grouped together in a single category called *Other*.

Course of Events on Day 2

The workshop kicked off with a discussion of the different centralities, the pathways between them and the future of the CCI building. Following the discussion the users moved on to working with the ColorTable. First they chose a view of the viaduct. S, the artist in the group, insisted on painting a mural of roses on the viaduct.

After that they meticulously positioned two pedestrian paths into the central panorama running through the park up to the chateau. They further augmented the scene with a bridge, football players, a string of grass chairs, and stairs. The scene they created also contained a number of sounds, like football players, children at play, people having conversations, and singing birds.

Then the focus shifted to the round parking lot adjacent to the CCI building and what to do with it. The Urban Sketcher was used to create a replica of the CCI building and to color it “bleu Klein”. Next they delineated the outline of the parking lot and assigned a new texture to it.

Afterwards the participants dealt with the issue of how to link the castle with the town. They debated about whether it should be accessible by car or by foot. This prompted a switch to the aerial view and they recreated existing conditions by placing a high traffic road onto the actual roads. Additionally, they placed another parking lot and a bus stop. Unfortunately, because of a misunderstanding regarding the scale of objects, the participants made the objects gigantic in size. They discovered their error after switching to the central panorama and corrected it. Subsequently, they decided on further changes in the park. Two ponds with water lilies replaced the old pond. They chose the see-through view to evaluate the placement of the different objects.

After another panorama was selected the focus shifted again to the parking lot. After some difficulty, they reduced it in size and added some green spaces. They also added a row of purple “cabanes”. This was followed by a string of green arches. E wanted to add a symbol but was unable to find one.

Prompted by the noise of a car, they started to experiment with sound aspects. They connected different sounds to the objects. Following that, they selected the fixed camera view and used the joystick to explore it. For a short time, they also used the mobile camera to check the positioning of different objects by directing the scout to different locations.

At the end of the workshop they concentrated on creating an entrance to the CCI building for pedestrians. With the Urban Sketcher a hole was cut into the blue building. The participants instructed one of the technical staff, who is more familiar with the Urban Sketcher, to make the changes for them. The workshop came to an end afterwards.

7 Data Collection

7.1 Video Material

The work with the ColorTable was documented in a number of different ways. Two members of the research team took pictures. The researchers wrote down special events and observations and interviewed the participants. The system of the ColorTable itself produced screenshots in intervals of two minutes and preserved the scenes the participants saved during their work.

The main part of the data collection is video material that was captured with two video cameras. A fixed camera was mounted on the side of the projection screen facing the participants and the ColorTable and a mobile camera was operated by one of the researchers and focused on significant events both around the table and on the screen.

Both views were synchronized and rendered into a split screen video. This has several advantages. The action can be viewed simultaneously from two different viewpoints. Watching first one and then the other view would be time consuming and would make it difficult to connect the events in the two videos. But with a split screen, one view might capture the actions of the participants, while the other shows the results of the intervention in the panorama on the projection screen. Furthermore, on some occasions actions that were occluded in one view could be seen in the other.

The video material for this analysis covers approximately two and a half hours. Both camera views include a time code, but the codes are not precisely in sync. There is a 32 second time difference between the two and the date of the mobile camera is set to the wrong year. The fixed camera was mounted upside down.

Using video observation opens up new possibilities for ethnographical studies. Two key advantages of video observation are *density of information* and *permanence*. Filming social interactions and behavior captures more information than any other observational method and the video material can be stored indefinitely. Other researchers can also view the data and conduct secondary analysis of the findings, thus potentially further validating interpretations and results. [41]

Digital video in conjunction with the flexibility provided by associated software augments the research process even further. The video stream can be viewed forwards and backwards at varying speeds ranging from scrolling through single frames to detect micro movements or rapid playback for quick overviews. [41]

Video management software, like The Observer XT [42] or Studiocode [43], also provides the functionality to tag or mark particular events like for example the occurrence of a gesture. These elements can be later extracted to create short video excerpts or combine similar events from different points in time. [41]

Even though video observation might seem to be the perfect tool for gathering behavioral data, it has its limitations. Video cameras have a limited field of view and some relevant behavior might take place out of range. Additionally, the spatial configuration of participants and/or objects in conjunction with a fixed camera angle can lead to occlusion. As explained above, during the workshops with the ColorTable two cameras where used to capture as much data as possible.

Introducing more cameras is problematic, because it can increase the participants' *reactivity*. Reactivity is the change in behavior due to the participants knowing that they are being observed. This effect might be mitigated by using very small or hidden cameras or by capturing more behavior. Only the material that was collected after the participants have moved beyond an initial phase of self-consciousness is then used for analysis. [44] Coleman [45] found when researching doctor-patient communication that even in the confidential and personal settings of medical consultations patients tended to forget that they were being filmed.

The participatory workshop lasted for two days and because the first day was also filmed, by the second day the participants showed no overt signs that their behavior was influenced by the presence of the cameras. By then they were also familiar with the research and technical staff that was present.

8 Qualitative Analysis

8.1 Framework

The research team used the collected data to perform an in depth qualitative analysis. The results were presented in the final report for the IPCity project. [5] The following eight research questions provided the roadmap for the qualitative analysis:

1. *In which ways do participants' gestures contribute to achieving an understanding of the site, explain and plan transformations of the site?*
2. *Which role do gestures play in the mapping of events in the RE [real environment] and the VE [virtual environment]?*
3. *Which body configurations, boundary crossings and collaborative modes does the MR-Tent enable?*
4. *In how far does the design of interactions contribute to participants' ability to construct a scene?*
5. *What are the specific social qualities of haptic engagement?*
6. *How do different types of visual and sound content enable participants to express and experience their ideas?*
7. *How does switching between scales and different representations of the site contribute to participants' understanding?*
8. *How does sound contribute to participants' engaging with a scene*? *[5]*, p. 1

When working with the ColorTable the participants create, discuss, or try to comprehend MR scenes, leading the research team to identify four different activity categories:

1. **Plan Intervention**
   Participants talk about the urban site, deliberate about possible interventions, and choose content.
2. **Perform Intervention**
   Participants place the virtual content into the scene and fine tune the positioning to fit their vision.
3. **Evaluate Result of Intervention**
   Participants asses their interventions and the impact they had on the scene. The assessment may lead to modifications or instigate new interventions.
4. **Understand MR Scene**
   Participants need time to orient themselves in the different virtual scenes and panoramas and to locate and comprehend their interventions.

This categorization scheme is also the basis for the quantitative analysis.

The activities were further broken down into four sets: gestures, body posture and gaze, object manipulations, and engaging with the scene. Figure 11 provides an overview of the different activities.

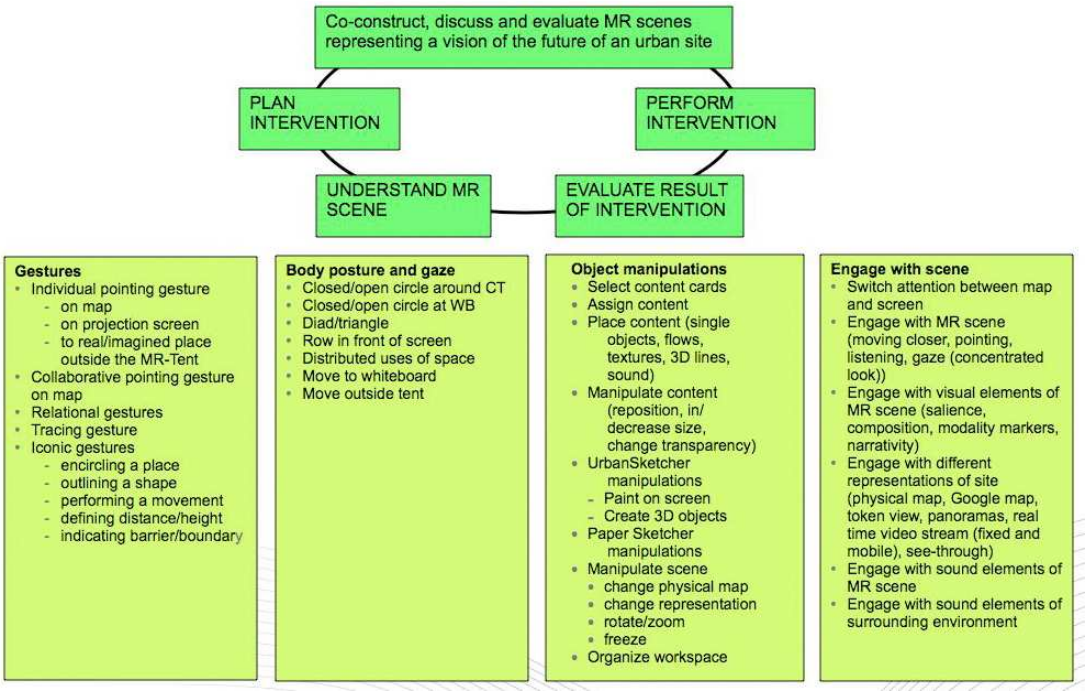


Figure 11: Activities [5]

The research team watched the video material numerous times and selected significant scenes based on the research questions and the different activities the participants exhibit in the setting of the MR-Tent (Figure 11). The selected scenes were than combined with the relevant screen shots, pictures, and interview data. These collaborative combinations of data allow the researchers to reach shared interpretations of their observations.

The results of this time consuming approach are presented in the form of "storyboards", which are a combination of descriptions of activities and sound, video stills, photos and screenshots as well as context information and parts of verbatim transcripts and are structured according to the observational categories. Figure 12 is an example of a storyboard, which describes how the participants manipulated a row of chairs.

The group has placed a series of chairs and they are looking for them in the panorama (P1) – G points with both hands

12.53:15 *We cannot see them since we have placed them closer to here and they are very small* (1) – V points them out on the screen – they start increasing the size

12:53:27 *Not too much, stop, we have to decrease again* – they look intently at the screen (2)

12:54:09 EV: *They are not really well positioned; we may have to push them to the other side?* - E and EV push the chairs to the other side of the path while the others look and direct (3), (4), Ch also helps

12:54:45 M: *They are between the Viosne and the … Well, are you satisfied?*

Increasing size of chairs – repositioning row of chairs

Figure 12: A storyboard [5]

8.2 Summary of Key Findings

Reference points for the quantitative analysis are the results of the qualitative analysis. The summary below has been taken directly from the final report. These findings also define specific aspects of the participants' interaction around the ColorTable.

Mapping between events in the RE and events in the VE is done through relational gestures (connecting events on the CT, the screen, and also the physical site), talking, gaze, and bodily orientation within the MR-Tent (Figure 1).

Figure 1:Understand MR scene: A participant tries to match places on the map with the MR scene: "This path here is this path over there" – points first on map (left) and then on screen (right)

Design guidelines:

Provide a sufficiently big map space 'within reach' for communicating urban issues with gestures that also supports the public visibility of action.

Consider screen size as an invitation to pointing and as an important aspect of immersion.

Participants collaboratively engage, enact and interact with small handles (content cards, command cards, barcode trays, and tokens) trough touching, holding, placing and moving these objects while discussing, reflecting or waiting. The objects stimulate different senses, creative use and support distributed attention (Figure 2).

Figure 2: Participant moves the sound token with the different material slowly coordinated with the speed of tracking from above (left). While discussing the flows across the bridge, a participant grasps a flow token gently beating with it on the map (middle). A participant places a command card on the configuration area while holding a second one (right).

Design guidelines:

Provide same conditions for each participant (equal accessibility to content, interface and projection) as common basis for discussions and visions.

Provide multiple handles using several small objects varying in form and material to support collaborative interactions and stimulate different senses and creative use.

Through haptic engagement with physical objects (content cards, tokens) participants signal forthcoming action in a way visible to all. Haptic lends an expressive dimension to their interactions. The physicality of the map invites grown practices of touching, pointing, and annotating that support the focused attention of all on an area of intervention (Figure 3).

Figure 3: Participants feel and touch the tokens while gently moving them (left). A participant's hand moving a content card in the middle of the table and withdrawing it immediately to grab a token, place it on the table and the content card on top of it. (middle and right)

Design guidelines:

Design possibilities to touch, feel and hold objects of various materials and forms.

Support multiple and creative ways for people to enact and interact with them.

Crossing MR boundaries forms an integral part of participants' understanding the urban site and their interventions. Being on site and sound play a large role in this process (Figure 4).

Figure 4: Switching to a real time camera view – fixed as well as mobile – requires another reassessment of the scene. Reality elements come into play, which stimulate boundary crossings. Participants are delighted to have real people mix into the scene (left). Participants discuss and create boundaries. They place flows and 3D lines of objects, lines and textures (right).

Design guidelines:

Explore different ways of relating real and virtual in a complex interface, including visual openings to the real site and sound.

Dynamic content (e.g. flows), 3D lines textures, and expressive content (e.g. content representing activities), the size and colour of objects influence the impression of 'realness' of a scene and help participants create narrative structures, insert borders and manipulate the salience of a scene. (Figure 5 and Figure 6)

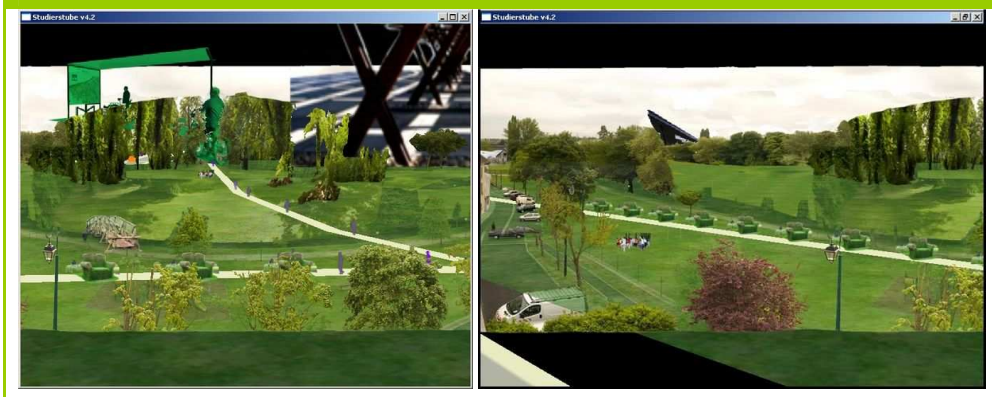


Figure 5: The two objects (parking for cars and bikes) that had been placed in the aerial view introduce an element of 'surrealism' into the MR scene (left). The participants are surprised, but content with the impact of what they perceive as symbolic interventions. The narrative elements of different MR scenes are strengthened by objects that represent activities (right).

Figure 6: Participants have just placed lines and textures that mark the parking zone they have planned (left). Coloured in blue the row of cabins pulls the blue building towards them, whilst the trees bind them even stronger together since they are encircling them and these geometrical forms step into the foreground; the green space with the cabins balances the CCI volume having an equal weight (right).

Design guidelines:

Provide dynamic and expressive content for participants to be able to create narrative structures and to compose a scene expressive to their ideas.

The availability of different forms of representations is a key feature of the MR-Tent; it

offers participants different possibilities for constructing, understanding, and evaluating MR scenes (Figure 7 and Figure 8).

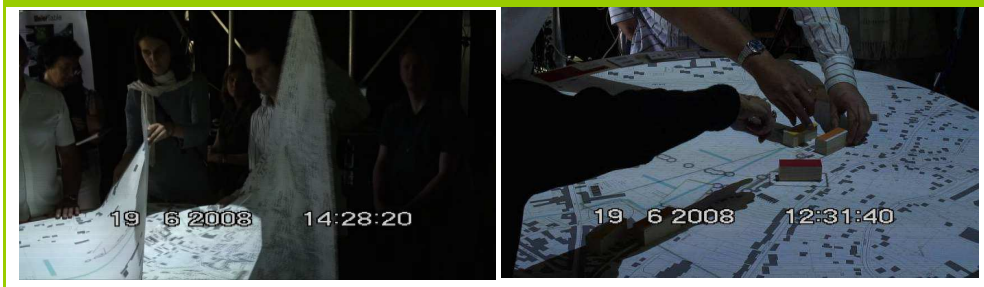


Figure 7: The physical map lends itself to planning and performing intervention at different scales (left). One participant directs action. Her pointing gesture is coupled with object manipulations performed by two other participants (right).

Figure 8: The panoramas are strongly edited views of the site with the advantage of providing a 360° view and space for interventions. It is mostly used for constructing scenes (left). Both real time video streams, fixed camera and scout, have a special 'realness' quality, which however makes the virtual elements stand out as 'model' or 'surreal' (right).

Design guidelines:

Provide a sufficient number of representations and scales that together cover the whole site, thereby enriching the opportunities for participants to realize interventions.

Permit that the same area or spot can be seen in different representations and from different viewpoints.

Sound is a key element of the participant experience, pervading what they discuss, see and do. Sound contributes to the blurring of MR boundaries; it strengthens immersion

into a MR scene; it contributes to the experience of spatial transformations; it evokes ambiences, thereby influencing action; particular interventions trigger engagement with sound.

There is the sound of birds. (11:56) Lots of laughter – all look outside

Ch: *But this not, this is not … magnifique, le chant du merle.* - E pointing outside: *So it is here* (as part of the panorama: in the bushes) *and Monsieur thinks it is there* (outside)! - B to G: *It is the sound of your small garden!* - B: *Where is the sound coming from? Is it here?* - 11:57:21 G: *This … du Hitchcock!* - G says this facing outside, the entire tent front is open, a sound of a motorbike starting up at the rue des Etannets driving past, softly breaking the silence.

In this scene we have three different sources of sound: the panorama sound (with a bird singing), real birds outside the tent, as well as a motorbike passing by. There is a blurring of MR boundaries – the bird sound could be part of the scene but also come from outside and it is associated with the imagined nearby garden of one of the participants.

Design guidelines:

Consider the importance of the real sound at different viewpoints (e.g. add panorama sound; transmit the sound of the AR view of the Scout) as increasing the sense of realness and immersion.

Create surround sound to strengthen immersion.

Figure 13: Summary of Key Findings [5] pp. 1 - 3

9 Quantitative Analysis

9.1 Framework

The qualitative analysis has shown that gestures are an essential part of working with the ColorTable. The goal of the quantitative analysis is to refine and augment the qualitative findings. Analyzing gestures in more detail might uncover new behavioral patterns and thus new insights, which can be used to further improve the experience of the users.

Pointing gestures have been identified as being very significant, because of the role they play in *mapping* objects, places, and events in the virtual environment to the physical map and the real world. Therefore the quantitative analysis has a special focus on relation gestures. Other relevant gesture types, for example iconic gestures, are also taken into account to obtain a comprehensive picture of the nonverbal communication around the ColorTable.

The overall framework of the quantitative analysis consists of the four activity categories: Plan Intervention, Perform Intervention, Evaluate Result of Intervention and Understand MR Scene. The gestures are categorized according to the specific visual features of the hand movements and then correlated to the activity category in which they occurred to see if specific patterns emerge.

Performing a gesture takes time, and during that period the performer is engaged with an external activity and/or his/her surroundings. To reflect this engagement in the analysis, the gestures are weighted according to their durations. For example, pointing at a landmark on the map very quickly suggest that the participant only used the map as a reference point very fleetingly and thus he or she is less engaged with the map. On

the other hand if the speaker keeps pointing at the map for a prolonged period of time, he or she is strongly focused on the map and wants to draw the attention of the other participants to the map and therefore all of the participants are more focused on the map. Since the analysis will draw conclusions from direction of the gestures and their ability to direct the focus of the participants, the duration needs to be taken into account.

Actions around the ColorTable, like using its functions and engaging with the haptic elements, are also documented and taken into consideration. They provide additional information about the context of a gesture and the characteristics of the activity categories and round off the analysis.

9.1.1 Gestures

Following the definition of gestures phrases described above, a gesture begins with a preparation phase, when the arm starts to move away from a resting position, culminates in a stroke, the phase that expresses meaning, and ends with a recovery phase, when the arm returns to the resting position. The same scheme was used to determine the beginning and end times for gestures in this analysis.

Gestures can also occur as part of a gesture unit. This means that a preparation phase is followed by two or more strokes and concludes with a recovery phase. In these cases the preparation phase was added to the first stroke and the recovery phase counted to the last stroke in the series.

The following types of gestures were chosen for analysis based on the on the qualitative analysis, except for beats, which were added to capture all occurring gestures.

Iconic Gestures

As described by McNeill [20], iconic gestures are pictorial representations of movements, objects, or shapes formed with hand movements in the air. They were

chosen because whenever somebody displays an iconic gesture, they express part of their imagination and ideas. No distinction was made between iconic and metaphoric gestures due to a language barrier.

	(7) Plan intervention E asks if it is possible to place a pedestrian path in a position elevated above the ground – he performs the gesture indicating the height level twice.

Figure 14: Iconic gesture [5]

Pointing

The purpose of pointing is to draw the attention of the interlocutor to points of interest in the surroundings. When collaborating around the ColorTable, the participants will need pointing to specify landmarks and areas for discussion, to orientate themselves on the map and in the MR scenes, and to coordinate their actions, for example when positioning tokens.

	(3) Perform intervention EV directs action: "To be able to walk towards 'Le Moulin de la Coulèvre', I would … in direction of the mill, voilà, like this maybe … ". Her pointing gesture is coupled with object manipulations performed by Ch and E.

(4) Plan intervention

Ch points at screen while explaining his plan: “There we have to reduce the parking … make a green space, which joins the other one on which we will place the cabanes, I think. We could redo a zone”.

Figure 15: Pointing [5]

Dual Pointing

In the environment of the ColorTable the urban site is represented in multiple forms. For example, a landmark like the chateau is marked on the physical map, but it can also be found in the corresponding panorama. Furthermore, another point of interest may be located just outside the MR-Tent. Dual pointing can be used to establish connections between the various locations.

(11) Understand MR scene:

J tries to match places on the map with the MR scene: “This path here is this path over there” – points first on map (left) and then on screen (right)

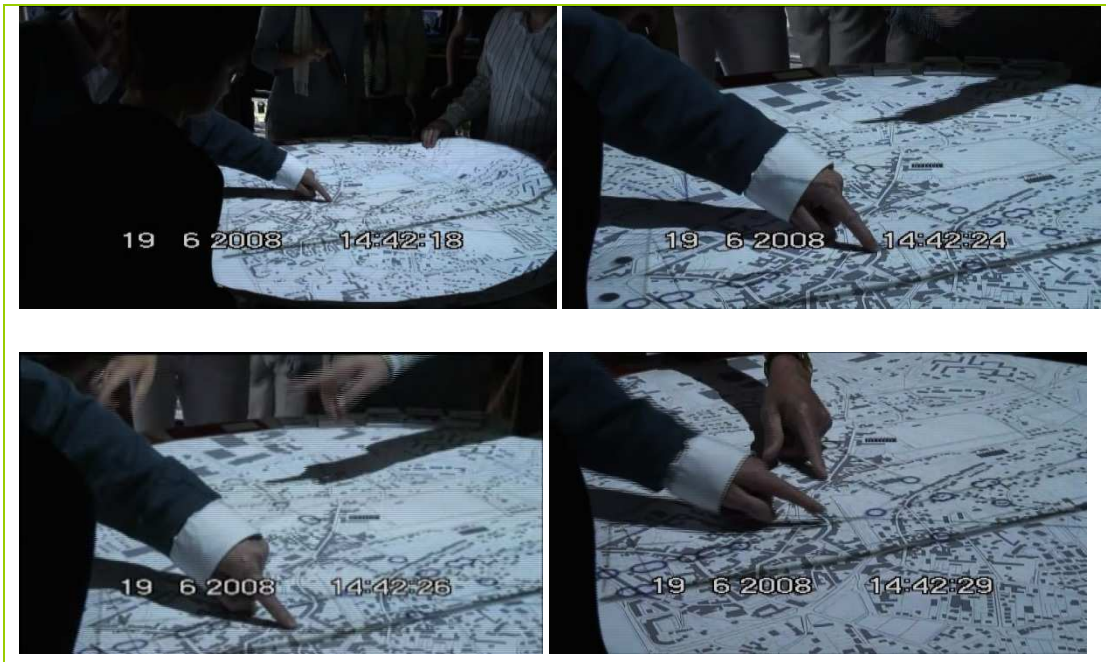
Figure 16: Dual pointing [5]

Dual pointing is documented in two different ways. Since they consist mostly of two consecutive pointing gestures, they are recorded as two instances of dual pointing with a different direction. In a separate database, the pointing gestures are combined into a single gesture with the shift in focus, like *map – projection screen*. Some of the dual

points also included a return to the starting point, for example *map – projection screen – map*.

Tracing

The physical map invites tracing gestures. They indicate how engaged the participants are with the urban site. Issues of access, flows and the course of streets and paths are often communicated through tracing gestures.



(10) Plan intervention

The group has changed map/scale to discuss connection between the CCI and Notre Dame. S and EV are discussing. S traces the road from CCI to Notre Dame with his finger: "North is there, the church is here … ". – "And then there was this idea of going up again … ". S moves his finger in direction of the church where it rests for a moment: "This is the cathedral, non?" (14:42:24). He moves al the way back and J's hand comes in.

Figure 17: Tracing

Encircling

The shape of an encircling gesture is often a circular motion with a pointed finger or all the fingers of a hand stretched out. Because encircling can encompass a larger are, it might indicate that a larger area is discussed rather than a specific location.

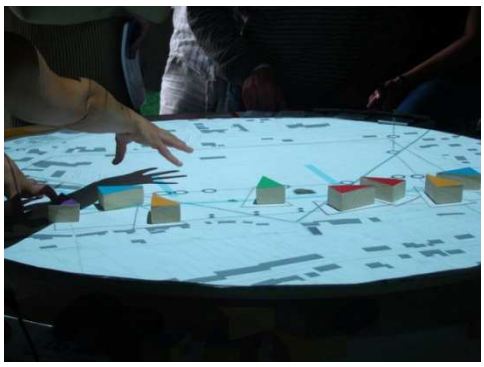


Figure 18: Encircling with an open Hand [5]

Beats

The function of beats is to emphasize and structure spoken language. They are of interest because beats allow conclusions about the engagement and the participation of a participant in the discussion.

Spatial Context

Additionally the directionality of each gesture is also noted to contextualize it with its surroundings in the MR-Tent. The following directions were observed during the video analysis:

- Air (i.e. the gesture space)
- Barcode
- Content card
- Flow
- Map
- Outside
- Another person

- Projection screen
- See through screen
- Table not map (area of the ColorTable that is not covered by the map)
- Token
- Whiteboard
- Other

Iconic gestures were mostly preformed in the air without any specific direction, but some of them were directed at the projection screen or the map. Due to their nature pointing gestures were focused on any of the directions with the exception of air. For the dual pointing gestures all the focuses were recorded. The physical maps are predominantly the locations of tracing and encircling. Because of their inherent function beats only occur in the air.

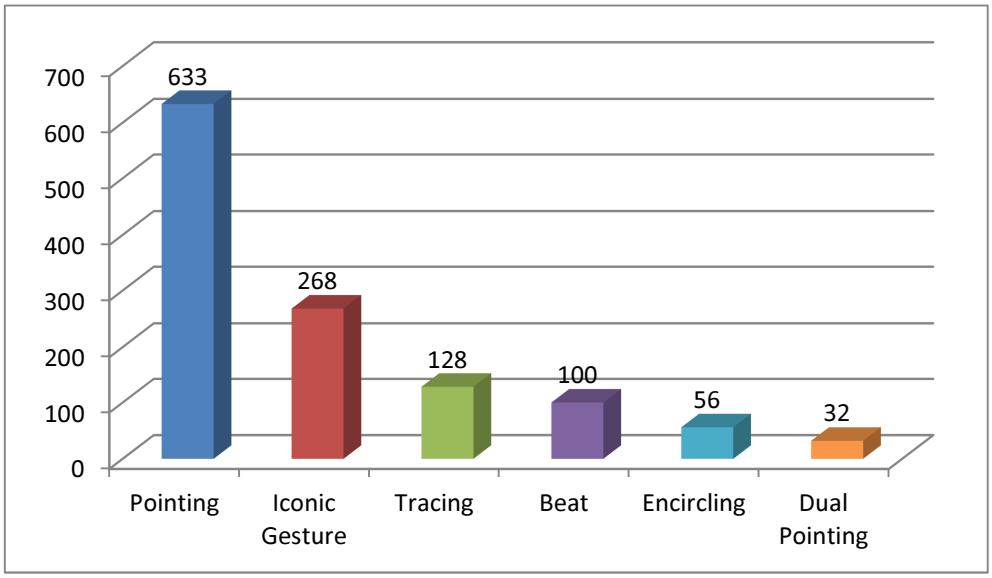


Figure 19: Absolut number of gestures

The grand total of gestures in the video material is 1444. As was indicated by the qualitative analysis, more than 50% are pointing gestures, because of the nature of the ColorTable and its role as a collaborative planning tool for urban projects.

9.1.2 Actions

In order to achieve a more comprehensive analysis of the behavior around the ColorTable, a number of actions connected to working directly with the different features and haptic parts of the ColorTable are also being taken into consideration. These actions can be grouped into two categories: object manipulations and scene manipulations.

9.1.2.1 Object Manipulations

To place virtual objects into a scene and adjust them to fit their vision, participants perform certain actions:

Holding

Technically, this action does not have anything to do with placing a virtual object, but it is of interest to observe how long and how often the participants hold tokens, content cards, or the barcode sheet in their hands.

Figure 20: A participant holding content cards for later use

Moving

Participants move tokens around on the map to position them in the virtual scene or to prompt tracking of a token. A differentiation is made between adjusting the positioning of a flow and a single virtual object.

Figure 21: A participant moving a token

Removing

The number of times a token or flow is removed from the map or a content card is taken away from the RFID reading field on the configuration board.

Figure 22: Participants removing tokens from the map

Unsure

Occasionally participants are unsure about their actions and undo them or start them and then stop half way through.

Figure 23: A participant contemplating a content card and then deciding against it

Handing over

One aspect of collaboration around the ColorTable is to hand content cards or tokens from one participant to the next.

Figure 24: Handing over a token [5]

Assigning

On the configuration board the participants can assign virtual content to the tokens.

Figure 25: A participant assigning a 2D object to a token [5]

Placing

Placing is the initial placement of tokens or flows on the map.

Figure 26: Placement of the starting position of a flow

Changing Object

Virtual objects have different attributes, like size, color, or distance from the floor, which the participants can change using a special subset of content cards.

Figure 27: Decreasing the size of a virtual object [5]

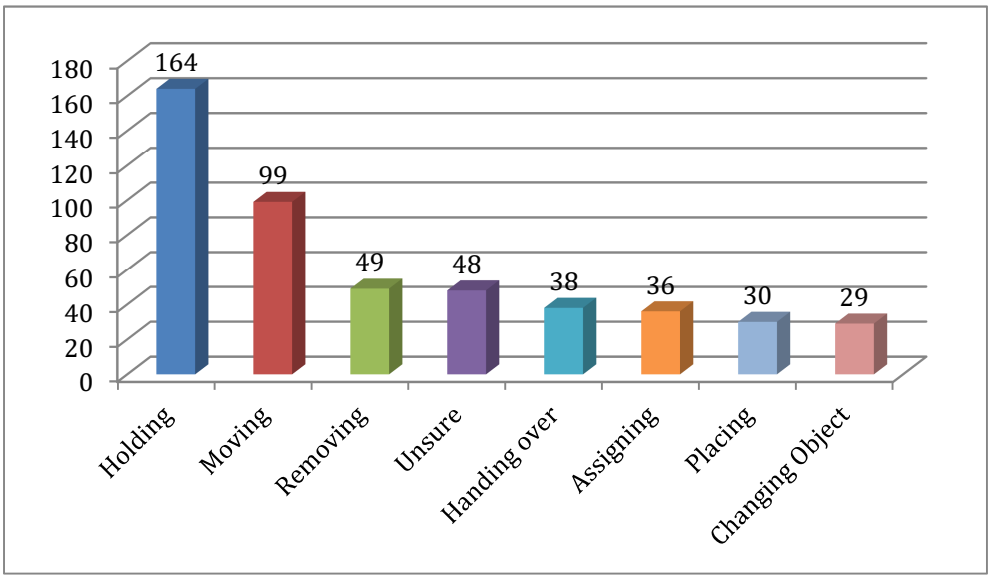


Figure 28: Absolut number of object manipulations

9.1.2.2 MR Scene Features

The ColorTable has a number of features that enable the user to explore and switch between different views:

Changing Map

Physical maps of the urban planning site are placed on top of each other on the ColorTable. The participants can choose between different scales or areas of the site

and place it on top. Due to the size of the maps the participants often work together to do this.

Figure 29: Participants changing the map together [5]

Changing Background

With a barcode reader participants can select different viewpoints on the map and switch to the corresponding panorama or activate the see-through view, fixed camera, or mobile camera.

Figure 30: Selecting a new viewpoint with the barcode reader

Rotate

A wheel under the ColorTable can be rotated to pan in the panorama and see different sections of it.

Zoom

The wheel also supports zooming into a scene.

Figure 31: Rotating and zooming is controlled by a small wooden wheel [5]

Freeze

The changes in a scene can be saved with the freeze function.

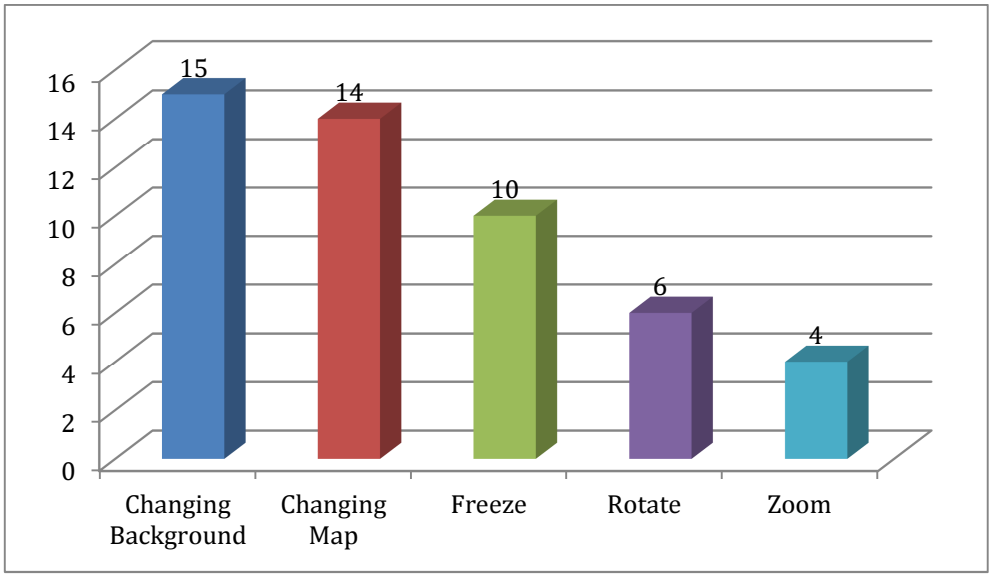


Figure 32: Absolut number of usage of MR scene features

9.1.3 Group Dynamics and Use of Space

Over the course of working with the ColorTable participants moved around the MR-Tent and formed groups of varying sizes. These group dynamics and use of space were also documented.

It was logged for each participant when he/she joined a new group and when he/she left it and which space the group occupied in the MR-Tent. Besides the ColorTable, groups gathered around the whiteboard, i.e. the content board, or in the free space between the table and the technical equipment. This space is referred to as *Other* in the analysis.

This data can be used to analyze changing patterns of collaboration and use of space. The analysis at hand will only look at the use of space in the MR-Tent. For an in-depth analysis of the use of space, collaboration, and body configurations in connection with the ColorTable see [5].

9.1.4 Activity Categories

The gestures, actions, and use of space were related to the activity categories by creating a timeline of the latter and then cross-referencing with the time stamps of the former.

Based on translated verbatim transcripts and the detailed qualitative analysis of interactions in [5] it could be determined how much time was spent in each activity category. Since participants frequently switch between categories and the transitions occur almost seamlessly, a meticulous analysis was necessary to truly capture their behavior. In the end, 137 intervals were detected in the video material and on average their duration was approximately one minute, but the durations range from three seconds to five minutes.

Figure 33 shows the chronological sequence of the activity categories and the number of participants that were actively involved at a given moment. The activity categories do not occur in a specific sequence. Even though the Plan Intervention category is almost always followed by performing the intervention, there are also instances were modifications of the virtual scene and objects are implemented based on the results of evaluating an intervention. Modifications of the virtual scene are always immediately evaluated. Often participants need to establish a new understanding of the MR scene after an intervention has been performed, but they also need to orient themselves in the virtual scene as a basis for planning an intervention. Overall the order and length of activity categories is determined by the flow of ideas of the participants, with occasional guidance from the moderator when the discussion gets too much off course.

The activity categories in a way constitute different modes of thinking. Thus, if gestures are truly a window into a person's mind, the types of gestures used and how they are used should be influenced by the activity category. This can be determined by taking a detailed look at the behavior in each category.

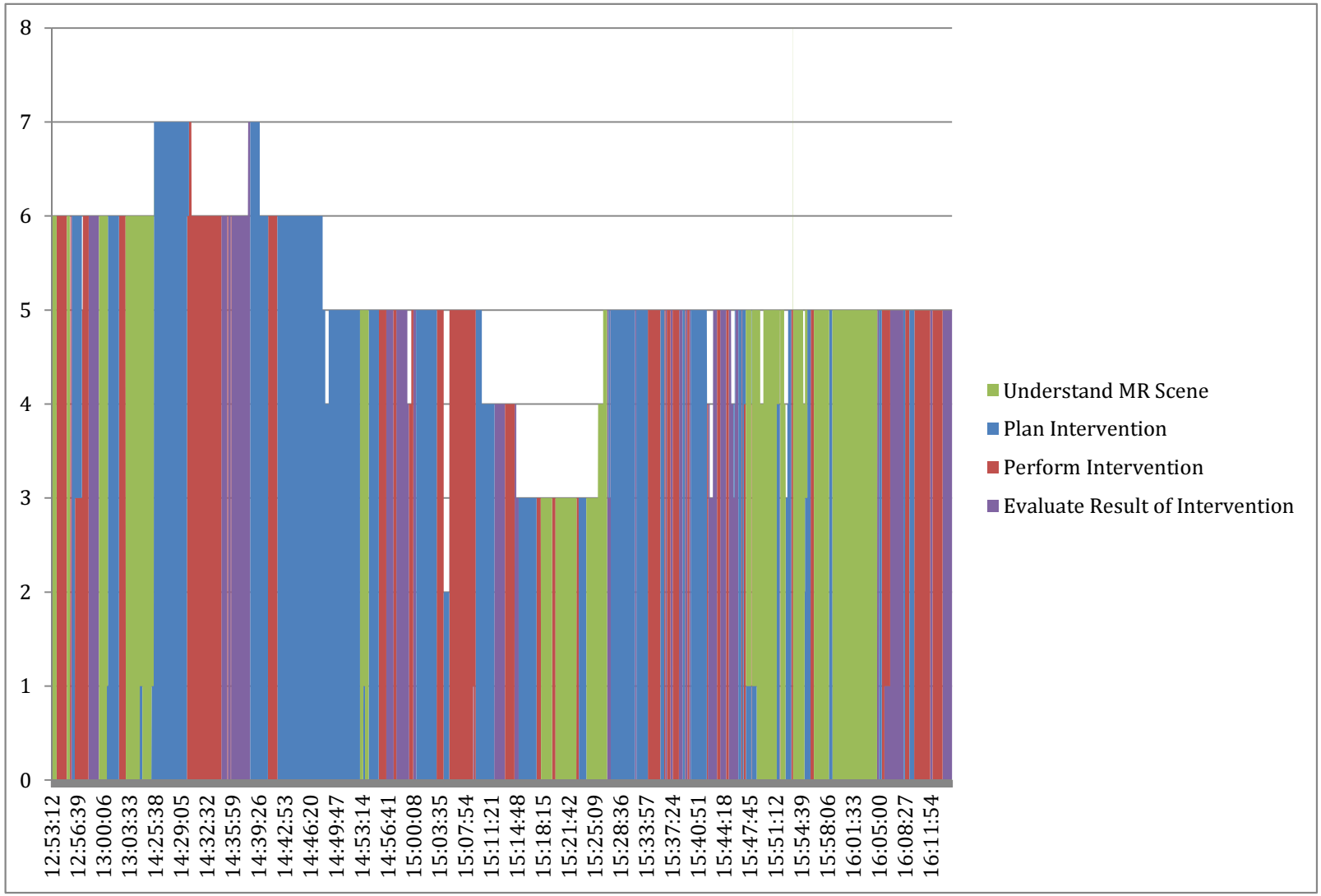


Figure 33: Timeline of the activity categories

9.1.5 Video Analysis

Two video cameras, one fixed and one mobile, were used to capture the activities around the ColorTable. The author of this thesis prepared the split screen video that was a source for the qualitative analysis and was also a part of the analysis sessions of the research team. These sessions afforded a deeper understanding of the ColorTable and valuable contributions to both the qualitative and the quantitative analysis.

The split screen video was then used to enumerate the specific gestures and actions of the participants. The relevant events only take seconds or less to perform. So in order to catch even minute movements, the video was viewed in Adobe Premiere Elements, because this program provides the functionality to scroll back and forth through the video in very fine increments. Furthermore, repeated viewings were necessary to detect the actions and gestures which occurred simultaneously.

Using the time code in the video, the beginning and end time were recorded for each action and gesture. This information is used to calculate the duration. Since the smallest unit in the time code is seconds, the events are always at least one second long, even though some of them do not last that long. For each gesture or action it is also noted who performed it and where a gesture was directed or what part of the ColorTable was manipulated during an action. To summarize, each data entry includes:

- Beginning time code
- End time code
- Duration
- Type of event
- Info field (where for gestures and what for actions)
- Person

9.2 Analysis of Activity Categories

9.2.1 Activity Categories

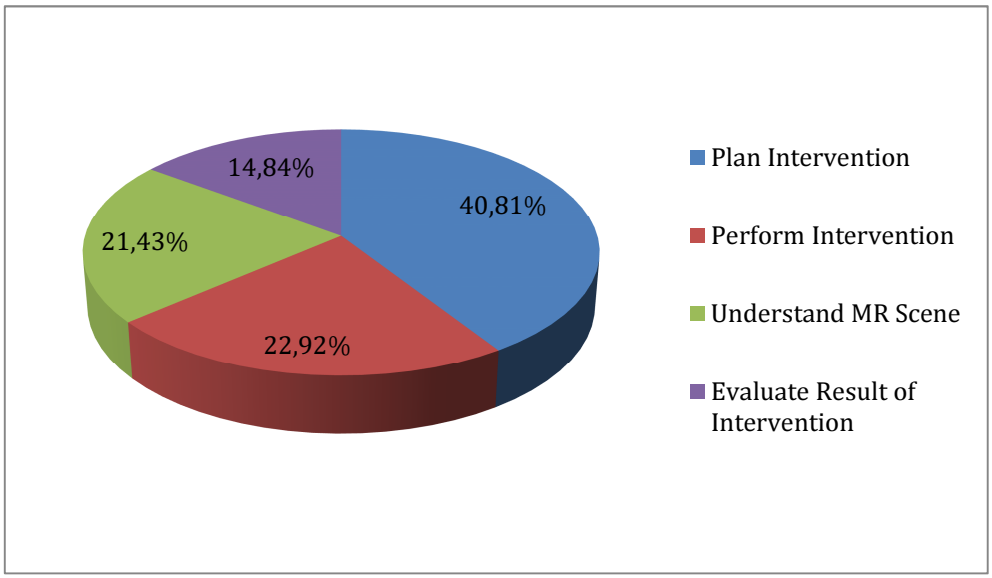


Figure 34: Time spent in each activity category

Most of the time was spent on planning interventions. The participants planned interventions for 40.81% of the time, roughly double the amount of time they needed to perform interventions or understand the MR scene, while only approximately 15% was spent on discussing the results of their interventions.

These numbers suggest that the participants discussed alternative options and planned their interventions in detail before transferring them into the MR scene, rather than trying out possible alternatives in the MR scenes. This does not mean that experimentation did not happen in the virtual realm. The participants for example evaluated varying colors, sizes, and placement of objects, but did not try out completely different possibilities. They developed their ideas outside of the MR scene and then focused on building an MR scene that reflected their visions. This would also explain why not much time was spent on evaluating the effects of their interventions, because

by the time the interventions were implemented they had already been discussed in detail during the Plan Intervention category.

That the participants needed approximately one fifth of the time to comprehend the MR scene might indicate that additional features are necessary to help with orientation. This should be considered with reservations, because it was often hard to differentiate between the participants evaluating a scene and trying to understand it, as those two categories tend to overlap.

9.2.2 Plan Intervention

General

During this activity category the participants develop and discuss possible interventions. Very often these discussions are centered around the map on the ColorTable. Thus the participants often start a new Plan Intervention category by choosing a scale of map and a viewpoint which best fits their current needs. After the intervention is fully planned out, they prepare to transfer their interventions into the virtual scene. Thus either the entire group, parts of the group, or individuals select the appropriate content cards from the white board. [5]

Use of Space

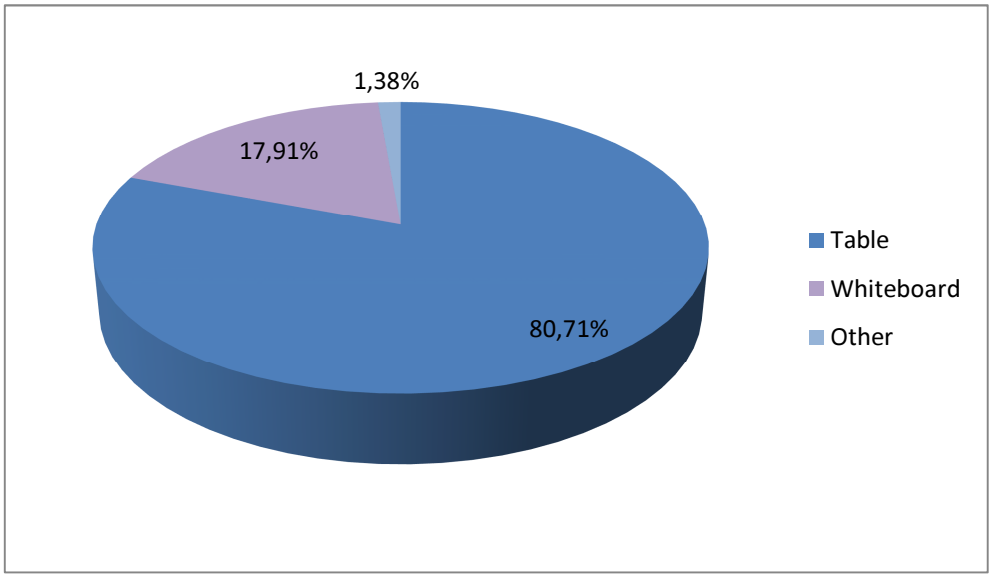


Figure 35: Use of space in the Plan Intervention category

Figure 35 shows that this behavior is reflected in the use of space within the MR-Tent during the Plan Intervention phases. The vast majority of time the participants are gathered around the ColorTable, but they also spend a significant amount of time in front of the whiteboard to choose content cards. This use of space distinguishes the Plan Intervention category from all the others. The time spent away from the ColorTable is negligible in the further categories. But we need to keep in mind that when the participants stand around the ColorTable, their focus can be either on the ColorTable and the map or the projection screen.

The qualitative analysis also found that the participants display special patterns of body configurations inside the environment of the MR-Tent. Three distinct patterns of use of space were identified:

Row/Line Formation

The participants stand in a row or line to have an unobstructed view onto the projection screen and to for example watch somebody painting with the Urban Sketcher.

Figure 36: Row/line formation [5]

Circle/Curve Formation

When collaborating and preforming an intervention, the participants most often form a tight circle or an open curve around the ColorTable. The open curve formation was also observed when the participants were watching an explanation from a slight distance without being directly involved with the table.

Figure 37: Closed circle around the table [5]

Figure 38: Open curve around the table [5]

Triangle formation

A small group discusses a specific point on the map. Their bodies are positioned in the shape of a triangle with the basis line of the triangle passing through the ColorTable.

Figure 39: Triangle formation [5]

Focus of Attention

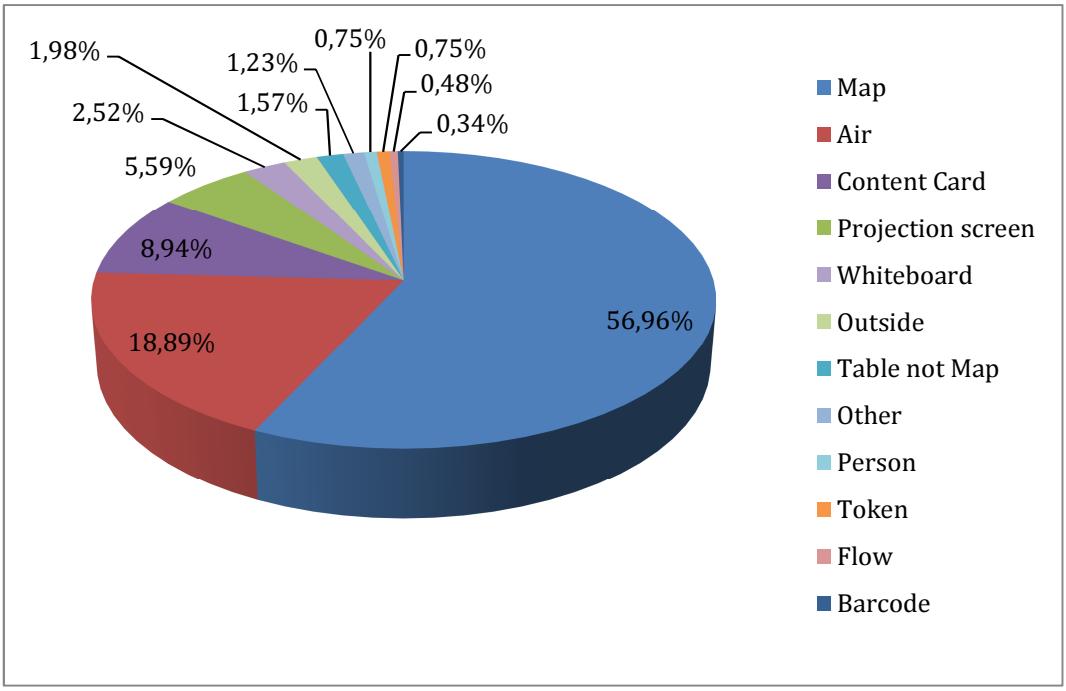


Figure 40: Focus of attention in the Plan Intervention category

Since gestures, especially pointing and tracing, are linked to their surroundings, one can use the directions of gestures (see Figure 40) to infer the focus of a conversation.

56.96% of the time, gestures were directed at or performed on the map. Thus the distribution of gestures supports the conclusion that the development of interventions relies heavily on the physical maps.

The projection screen, on the other hand, was only referred to 6.09% of the time during the Plan Intervention phase, while 75.32% were used to perform gestures in the proximity of the map and in the air. These numbers confirm that the participants planned and discussed their visions outside of the virtual realm. Rather than trying out a number of options by placing different objects in the virtual scene, they examine different ideas by talking about them and using the map as a reference to the urban planning site.

Both the content cards and the whiteboard are connected to selecting virtual content. These two sections combined make up only roughly 11%. Taking into account that 17.91% were spent in front of the whiteboard, this means that the participants did not take a lot of time to choose contents, leading to the assumption that most of the time they picked the contents that best represented their ideas, rather than the content cards being a source of inspiration. These observations lead to the question of how people used gestures to share their ideas.

Gestures

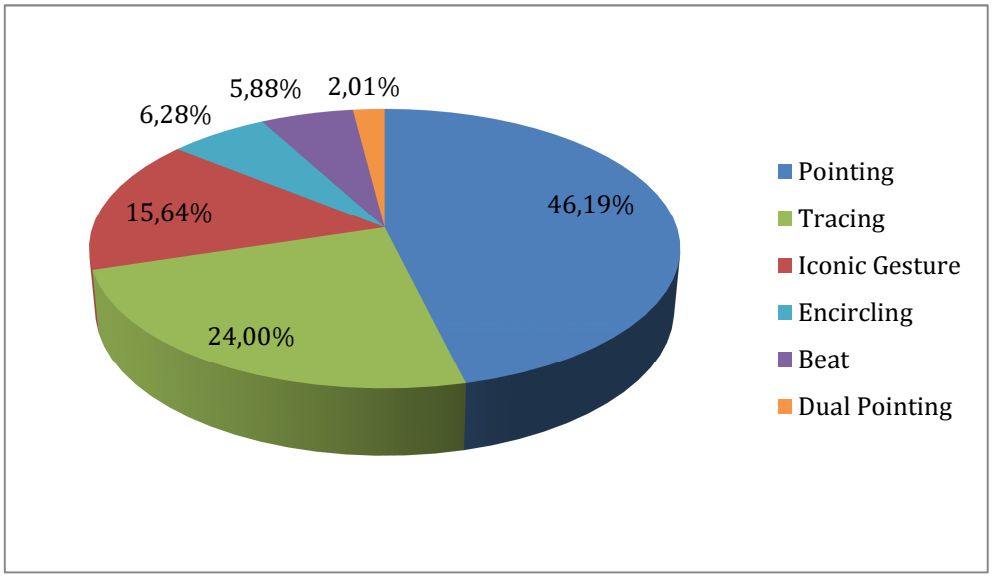


Figure 41: Distribution of gestures in the Plan Intervention category

Pointing is the most common gesture in all of the action categories. During the planning of an intervention it is used to help participants to orient themselves on the map, identify points of interest and, in a more metaphorical sense, to establish a connection between their ideas and the urban planning site.

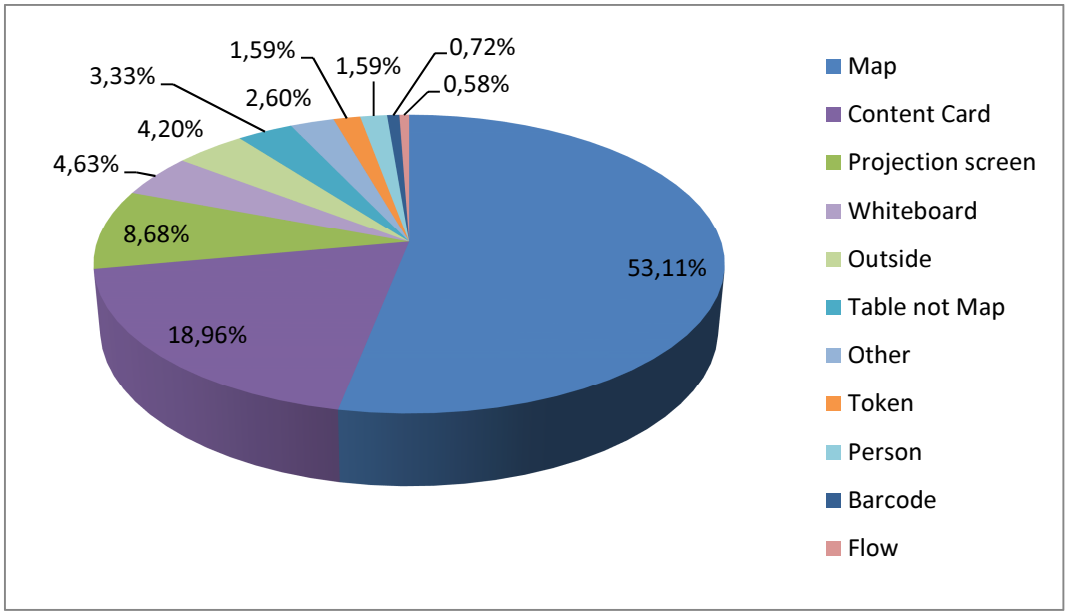


Figure 42: Focus of the pointing gestures in the Plan Intervention category

More than half of the pointing gestures were directed at the map. The second biggest category is content cards and, again, the projection screen is only the target in 8.68% of the cases. Furthermore, it is interesting to note that only a small number of encircling gestures were observed in comparison to the large amount of pointing gestures. This might suggest that the participants focused their intervention more on distinct locations and did not pursue changes of larger areas. This does not imply that the small changes were not coordinated to have an impact on a bigger scale. But the information on the meaning of encircling in the context of the ColorTable is not enough to make definitive assumptions.

Tracing is closely related to pointing. Like pointing, tracing helps the participants to get a bearing on the map by for example identifying existing roads or connections. Participants also trace new roads and walkways to explore how connections can be established and impassable areas can be made accessible. Furthermore, tracing gestures play an important part in determining the precise course of a road or flow.

96.94% of all tracing gestures are carried out on the map. An extra amount of focus on the map is required to execute a meaningful tracing gesture and sometimes the finger

glides directly on the surface of the map adding a tactile component to the gesture. This suggests that tracing gestures might play an important role when it comes to the participants getting a deeper understanding of the urban planning site and developing strong links to it.

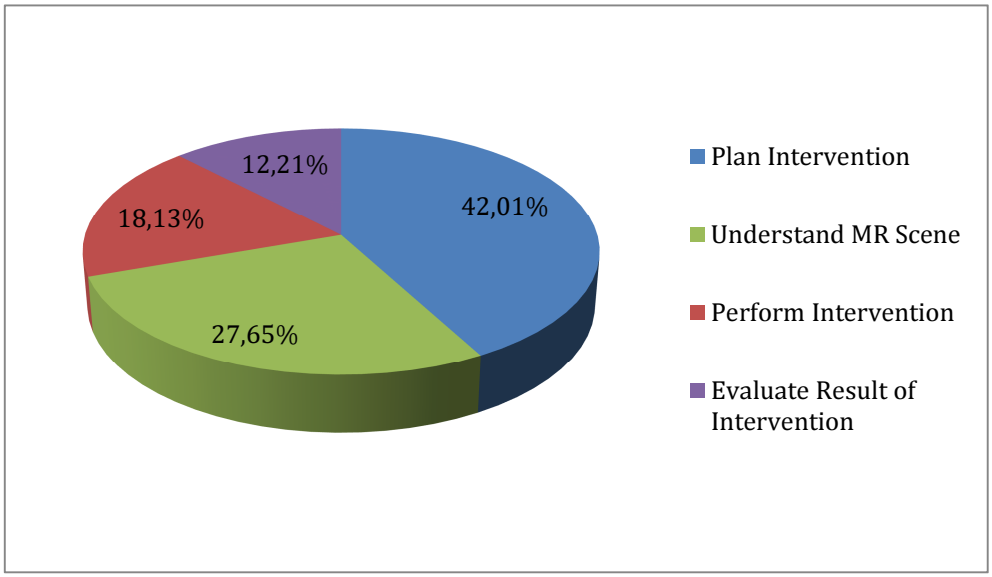


Figure 43: Distribution of iconic gestures in the activity categories

42.01% of the iconic gestures happen in the Plan Intervention category, but interestingly, the time spent with iconic gestures in the category itself only amounts to 15.64%. Participants might use them to express the visual components of their ideas. Furthermore, over all the categories combined iconic gestures make up only roughly 20%.

This might indicate that the focus is more on what kind of intervention they want to perform instead of what an intervention will look like in detail. A simple rhetorical example can illustrate this, namely, that the idea of placing a pond and what might represent that pond is more relevant to them than the exact shape of the pond.

Object Manipulations

During the Plan Intervention phase the number of active object manipulations, which result in actual changes in the MR scene, is insignificant. Participants spent 95.59% on holding content cards, tokens, and the barcode sheets. This result substantiates that participants choose content and prepare to perform interventions, but holding these physical objects over long periods is also very prevalent in the other activity categories. The shape and size of the tokens and content cards were specifically designed to easily fit in one hand and holding them does not interfere with other activities.

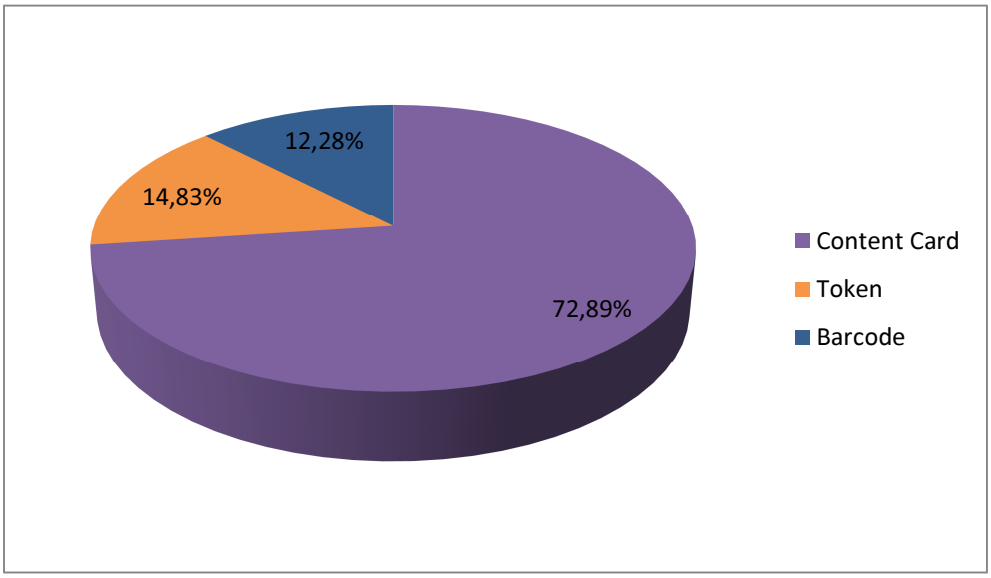


Figure 44: Holding physical objects during the entire course of the workshop

During the entire course of the workshop participants liked to keep content cards in their hands, while tokens and the barcode sheet were picked up when an intervention was more imminent. Content cards are also held to be at hand when needed, but content can be assigned to a token randomly, while a content card stands for a specific idea. This indicates that content cards serve as a handle on ideas and also function as a reminder.

Scene Features

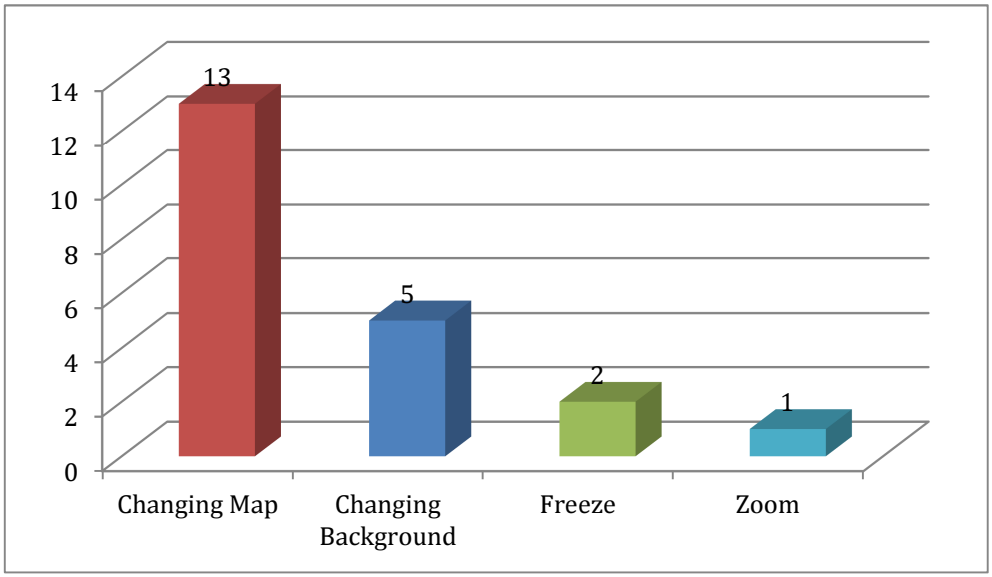


Figure 45: Usage of MR scene features in the Plan Intervention category

The physical map is the focal point of the discussions during the Plan Intervention phase. Thus, participants put careful consideration into choosing a map to work with, and changing maps is basically done only in this stage. The count of changing maps is misleading because it does not count the map changes as such, but all the participants that were in some way involved, from lifting the maps to slight adjustments of the placement on the table. On average, two to three participants switched maps together. The final step in the workflow was to select a point of view with the barcode reader. The lack of zoom and rotate activities and the low number of background changes further strengthen the observation that the participants developed their ideas on the physical map rather than exploring the virtual scene for inspiration.

The map is also within arm's reach. The participants can touch it. Drawing the attention of all the participants to a specific point can be accomplished much easier on the map than on the projection screen. Due to the setup within the MR-Tent the table is kind of a barrier between the participants and the projection screen, but furthermore the closer the participants would get to the projection screen the more their shadows would occlude the scene.

Taking everything into account it might be of interest to redesign the ColorTable and the workshops in such a way that the planning process is transferred into the MR scene. This step might benefit from the content becoming more flexible and the Paper Sketching system [5], which enables the users to incorporate their on-the-fly sketches into the virtual scene and which was used in a later workshop, is a move in that direction. But at the same time one would have to observe if the design process would even benefit from a shift into the MR scenes.

9.2.3 Perform Intervention

General

After the participants have developed their ideas, they transition into the Perform Intervention activity category. The ColorTable system offers various features that enable the participants to transfer their vision into the MR scenes. Objects, 3D lines and flows can be placed by positioning the colored tokens on the map. Furthermore the properties of objects, like color, size, or offset from the ground can be modified. The immersive nature of the scene can be enhanced by introducing sound into it. New geometric objects can be created with the Urban Sketcher, but this feature is difficult to use and thus can become very time-consuming. [5]

Use of Space

Since the content cards have already been selected at the Plan Intervention stage and the tokens are conveniently stored under the table top, the participants can perform their interventions without having to move away from the table. They gather around the table in a semicircle which opens up to the projection screen, so that everybody has an unobstructed view on the map and the screen.

Focus of Attention

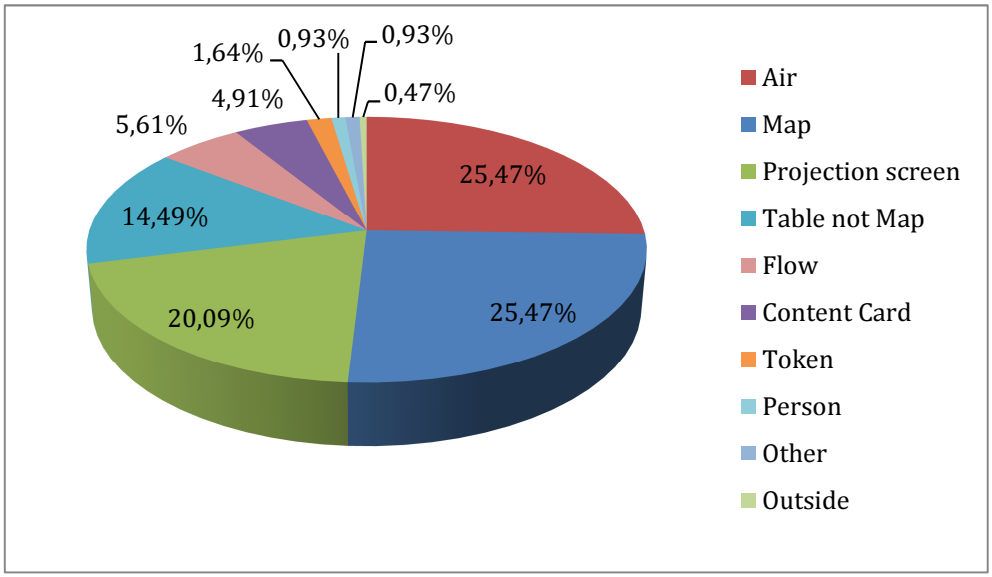


Figure 46: Focus of attention in the Perform Intervention category

The directionality of the gestures shows that the attention has shifted from the strong focus on the map during the Plan Intervention phase to include the projection screen and the configuration section and the storage area of the table (the last two subsumed in *Table not Map*). This is even more strongly reflected in the pointing gestures the participants used to draw attention to these three parts of the ColorTable setup. More than 75% of the pointing gestures were directed at these three locations.

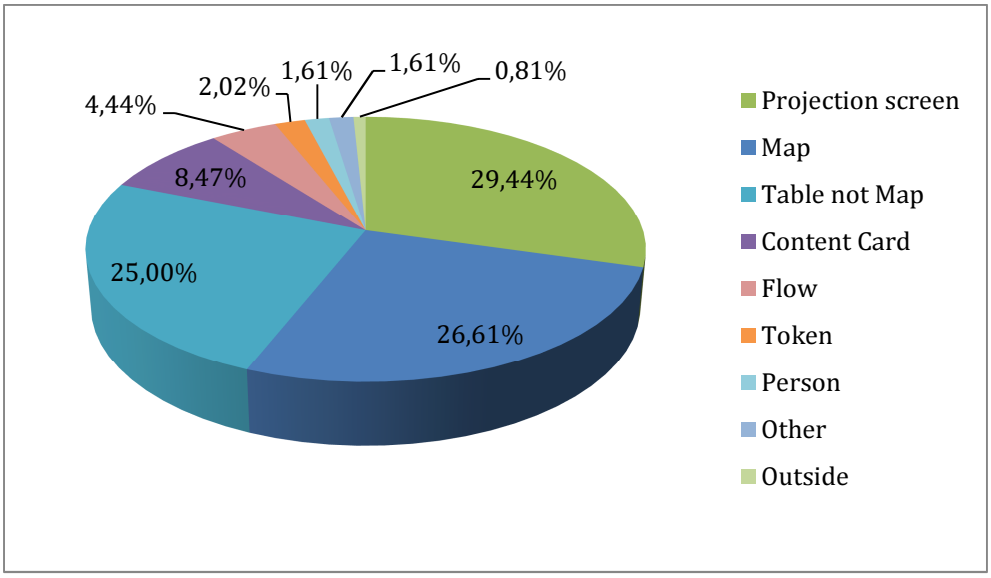


Figure 47: Focus of pointing gestures in the Perform Intervention category

The participants take certain steps when performing an intervention. These steps do not have to follow a strict order. First they pick a colored token and assign content to it or vice versa. These actions involve the configuration and storage sections of the table. Then the tokens are placed on the map to position the object in the virtual scene. The participants constantly switch their attention between the map and the projection screen in order to check the effect of their actions and to fine tune the positioning. Placing a flow often involves two people and each is placing one of the end points. Getting the line of the flow to fit their vision can be a tricky process.

Changing the attributes is also an important part of performing an intervention. The change is triggered by placing the corresponding card in the configuration area, which starts an incremental change of the object. The attention switches immediately to the projection screen to track the progress and to make sure to remove the card at the right time to stop the change.

The distribution of the pointing gestures stems directly from the different work steps, because each of them requires to a certain degree a different focus of attention. Assigning and modifying content involves the *Table not Map* areas. Placing the token is done on the map and the result is checked on the screen. That the time spent on

pointing is almost the same for the three areas suggests that the participants have understood and mastered the steps necessary to perform an intervention. If they had a significant problem with any of them that would probably result in a substantial shift of attention to one area. This assumption also takes into account that only 22.92% of the time was spent in the Perform Intervention phase. Therefore not a lot of time was needed to perform an intervention in general and the possibility that all of the steps were equally problematic can thus be excluded.

Gestures

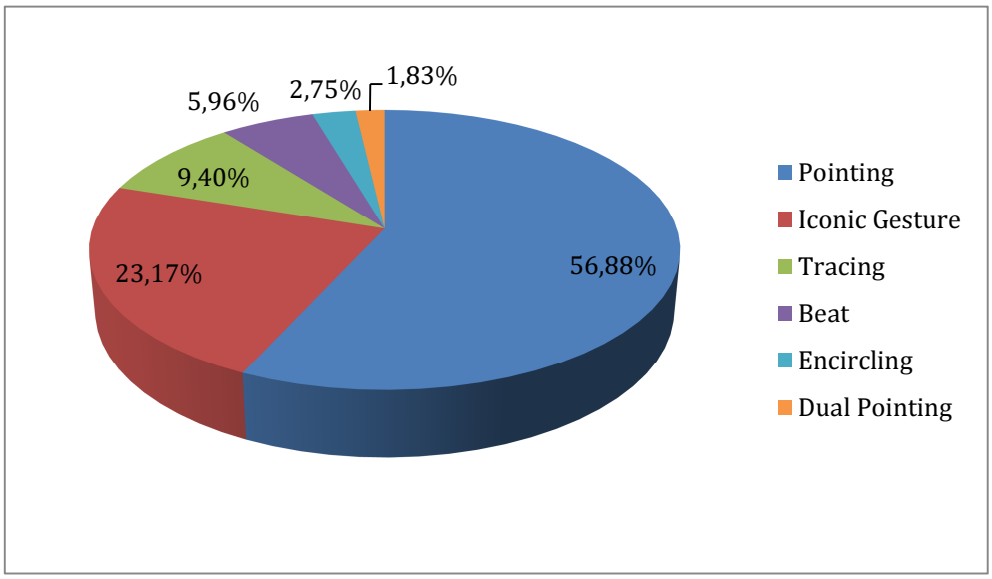


Figure 48: Distribution of gestures in the Perform Intervention category

As in all the other activity categories the majority of meaningful hand movements are pointing gestures, but at this stage pointing is used to coordinate and organize the highly collaborative process of performing an urban intervention. For example, often one person would assign the content, while another would place the corresponding token on the map. Placing flows was also almost exclusively done by two people. In order for this collaborations to lead to the desired outcome the participants relayed on pointing. Pointing was used to identify and/or select one of the colored rectangles or content cards on the configuration area or to ask somebody to hand over a specific

token or content card. The participants would also point at the positions where tokens should be placed or moved to. Finding and drawing attention to the results of the intervention in the virtual scene was also done with pointing.

The percentage of iconic gestures is slightly higher in the Perform Intervention category. Participants express different sizes, spacing between different objects and within a flow, and to a certain degree varying shapes of objects. These aspects become somewhat more relevant when the interventions are transferred into the MR scene.

Tracing gestures delineated the course of flows and roads to aid with placing them correctly. Their layout has already been decided in the Plan Intervention phase and the projection from above superimposes a line that represents the flow on the map, thus tracing gestures are much less relevant at this stage. The participants are also very focused on the exact points of their interventions; therefore encircling gestures are almost nonexistent.

Overall the participants are highly involved with the application functions of the ColorTable during the Perform Intervention category. They know what they want to do and the design of the ColorTable makes it easy to translate their ideas into the MR scene. They use pointing gestures to coordinate their efforts, but they have no need for lengthy explanations that might require more gestures. Thus, only 15.23% of all gestures happened in this category.

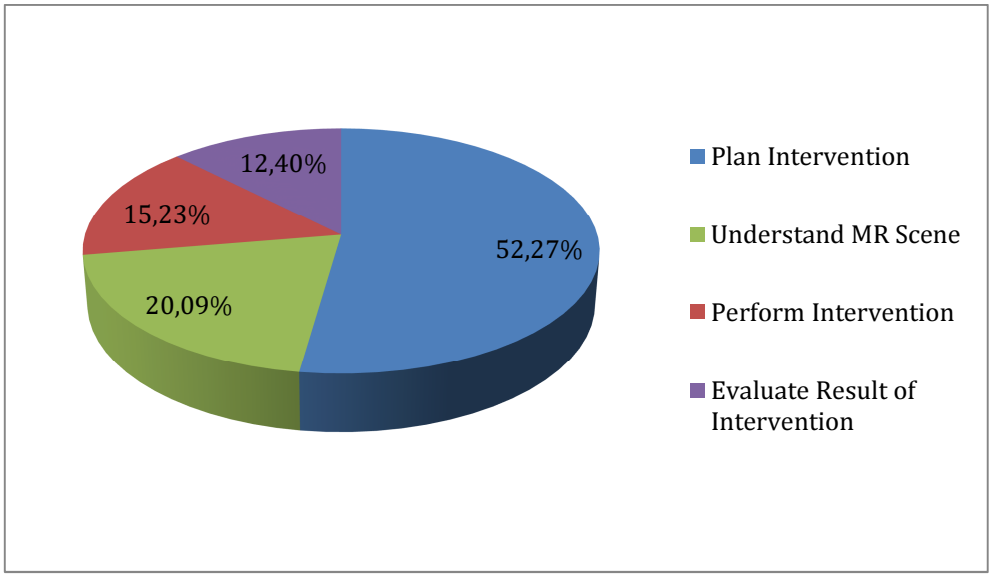


Figure 49: Distribution of gestures in the activity categories

Object Manipulations

This analysis of the use of application functions includes observations from [1].

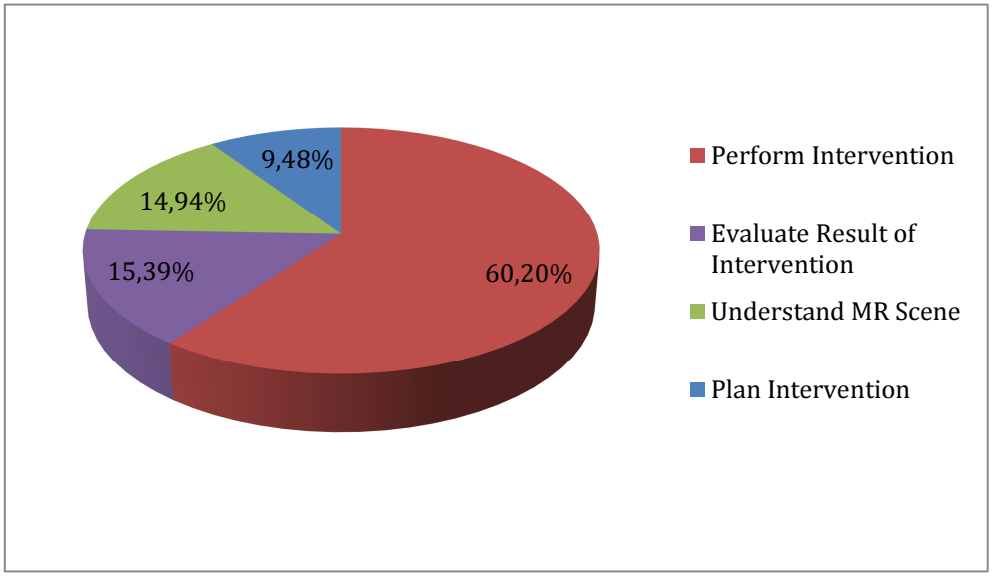


Figure 50: Distribution of actions in the activity categories

Rather than by gestures, the Perform Intervention category is characterized by the fact that it is the most active category when it comes to the interactions with the

ColorTable. 60.20% of action time takes place in this phase (*Holding* is not taken into account). The participants are very focused on modifying the virtual scene to fit their visions and make full use of the technical possibilities provided by the ColorTable.

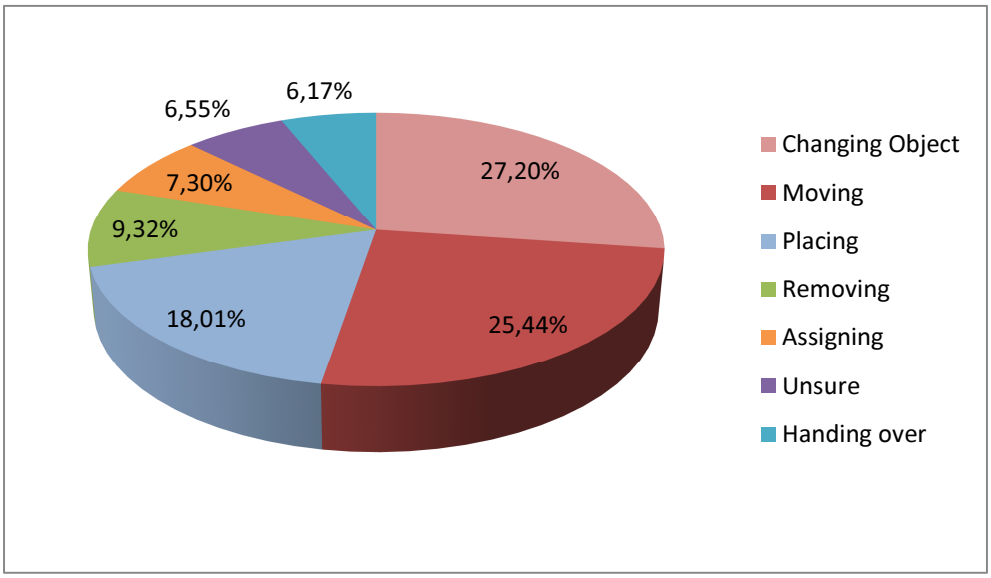


Figure 51: Distribution of actions in the Perform Intervention category

Figure 51 depicts the distribution of actions in the Perform Intervention category. The participants spent a considerable percentage of time on changing the virtual objects. Compared to the other interactions, modifying the attributes of objects is relatively time consuming because it is an incremental process and the participants have to pay close attention that they do not pass their intended setting. Increasing and decreasing the scale of an object was done to fit the object into the panorama, but the participants were not always very strict about the size of objects. For example, due to performing an intervention in an aerial view the scale of a parking structure was gigantic in the panorama. The participants decided to leave it that way, because for them the size represented the impact this intervention would have on the area.

Color was often chosen based on aesthetic criteria. The participants would run through the different color choices a number of times and experiment with them before they settled on one. Changing the offset was done to adjust the perspective of a 2D billboard

to the panorama. Modifying the spacing of objects in a row does not only influence the distance between objects, but also their number. Thus, the participants could gouge how many objects, like for example cabanes, were needed and practical.

Placing tokens on the physical map introduces virtual content into the scene. The 18.01% include the time from picking up the token to releasing it again on the map. If the participant had been holding the token for a while, the start of placing would be the beginning of movement towards the map. The participants took their time to make certain the initial placement was close to the point they wanted it to be. Objects are introduced not only to represent a physical entity, but also to add further characteristics to another object or to mark an area were a certain activity might take place. For example, stairs to add further accessibility to a connection or tai-chi practitioners are added to an area. This was also done to add a sense of tranquility to that area.

Adding a ground texture to an area is likewise done by placing a round token. Participants used textures to mark zones for different usages. A parking lot, for example, was added by assigning a grey texture. A grassy area was represented with green.

The boarders of these areas were defined by placing roads and flows. The placement of a road or flow requires two tokens which specify the endpoints. Usually, two participants work together to adjust the endpoints to their needs. Besides functioning as outlines, they established connections between urban locations or delineate routes to cross a specific area. The participant also added flows of pedestrians or cars to existing roads. Introducing animated elements into the scenario made it more realistic and engaging.

Once the virtual object appears in the panorama after the corresponding token has been placed on the map, the participants start to fine-tune the position of the tokens to get the best results in the panorama. Roughly one fourth of the action time was used for moving tangible pieces on the table, thus getting the alignment with the panorama just right has a high priority for the participants. However, the participants also moved

tokens to facilitate the tracking process which could be faulty at times or to see which colored token was assigned to which virtual object.

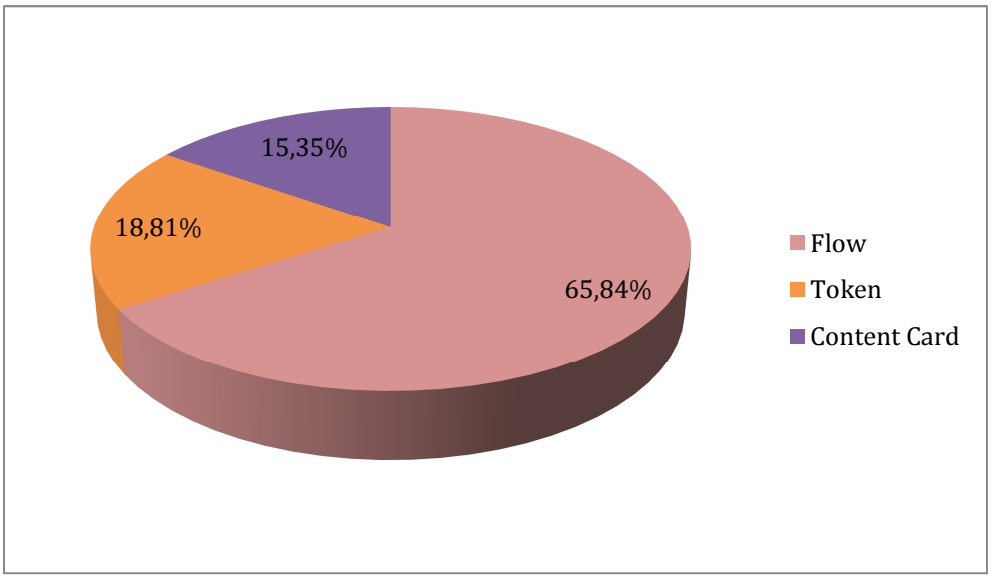


Figure 52: Moving physical objects in the Perform Intervention category

The majority of movements dealt with tokens that were paired together to create flows or roads. This is an indication that setting the course of a flow is important to the participants, because they are willing to spend a significant amount of time on it, and that at the same time the way of manipulating flows might be too complicated. Therefore, finding a simpler and more efficient way to make flows might be a valuable improvement.

The content cards were primarily moved to organize the workspace. For example, cards that had already been used were put into stacks to have them out of the way or cards that were needed for the next intervention would be moved to have them closer at hand. Part of the organizational activities was also to remove tokens from the table and content cards from the assigning areas of the configuration board after a scenario had been frozen.

The participants were quick with assigning content to tokens. Only 7.30% of the active time was needed to do it. They had already chosen and prepared the content cards

during the Plan Intervention phase and this relatively short period means that they implemented their vision as they had planned and did not experiment with completely new ideas.

The thorough preparation also made the participants very sure about their actions. They merely hesitated 6.55% of the time. This means that the participants had a good understanding of how to operate the different application functions of the ColorTable and working with it was a fluent process.

Scene Features

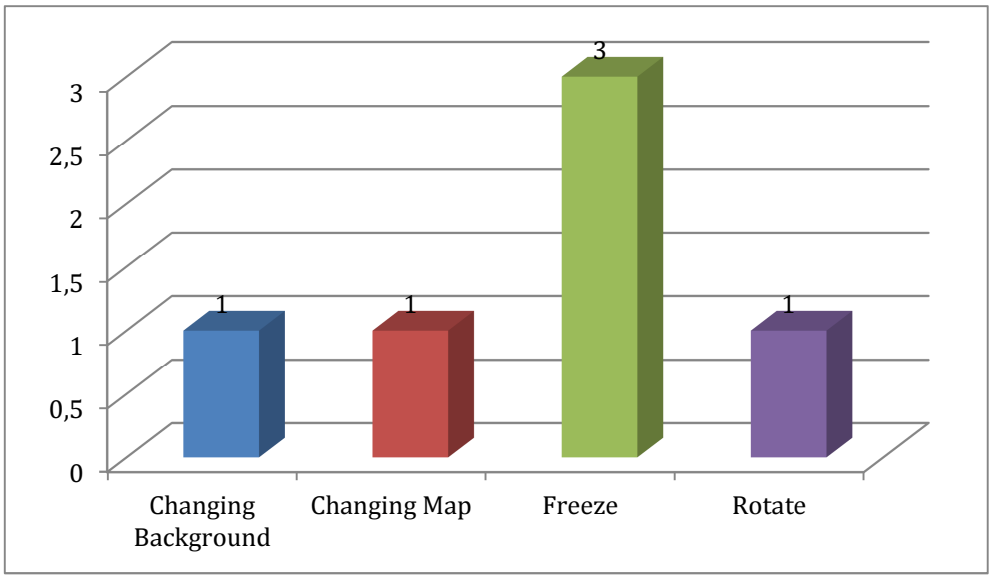


Figure 53: Usage of MR scene features in the Perform Intervention category

Since the map and background had already been chosen in the Plan Intervention phase, the participants needed to change the map and background only on one occasion to be able to carry out their interventions. Freezing a scene during the Perform Intervention phase was done because the map had become too cluttered and the participants saved their changes so that they could remove the tokens and make room for further interventions. When the participants were searching for virtual objects, this was considered to be part of the Understand MR Scene phase and evaluating the result of

their intervention is also a separate category. The view on the screen is not essential for the act of performing an intervention as such.

9.2.4 Evaluate Result of Intervention

General

Evaluate Result of Intervention is the smallest category with only 14.84% of the observation period. Again this low number can be ascribed to the comprehensive discussion of interventions in the Plan Intervention phase and the fact that the participants do not experiment very much with different options in the virtual realm. Thus, the assessments of interventions are mostly very brief and it was not easy to detect them.

But there were some extensive discussions surrounding some controversial issues, where the participants needed more time to find a common ground. One such example is the area of the CCI building. The participants had concerns about the size of parking, the placement of a new building and the cabanes. [5]

Evaluating interventions can trigger new ideas or bring another issue to the foreground and thus resulting in a switch of focus.

Use of Space

Looking at the scene that has been created is an integral part of the evaluation process. Hence, the participants are arranged around the ColorTable and oriented towards the projection screen. That the whiteboard is not part of the discussion suggests that they are not weighing completely different interventions and ideas against each other, but that the focus is on the attributes of the interventions, like size, spacing of the objects, color, or the precise position.

Focus of Attention

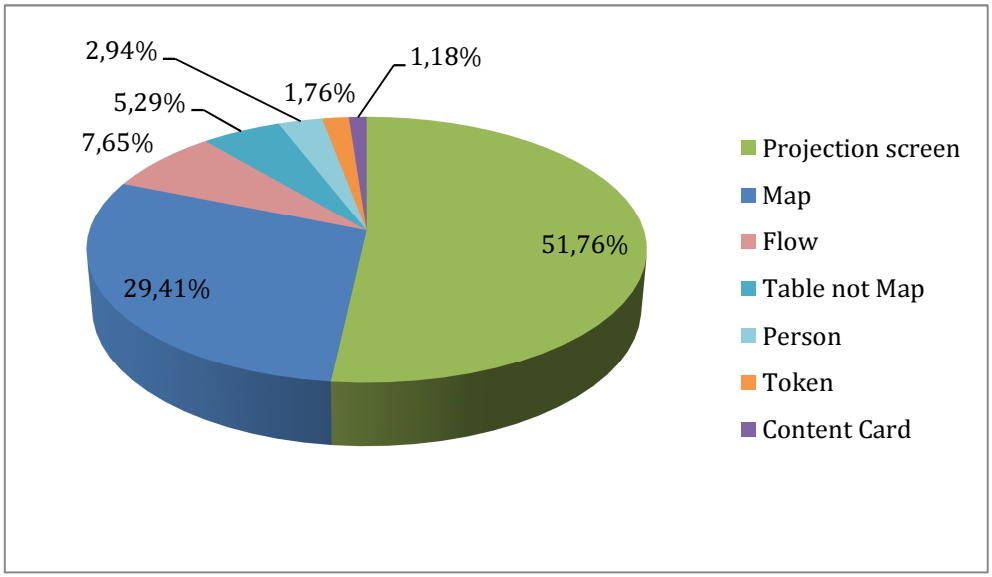


Figure 54: Focus of attention in the Evaluate Result of Intervention category

During the evaluation process the attention is strongly focused on the MR scene, because the participants discuss the impact of an intervention on the overall scenario. The 29.41% of the time the participants referred to the map was mostly done to locate certain interventions and on a few occasions to introduce new ideas.

Only 7.65% of the pointing gestures were directed at flows, which leads to the conclusion that the participants were very certain about the connections and boundaries they had created. Since those aspects pertain to the entirety of the urban planning site, this might mean that the participants had a clear overall vision for the site and were focusing on interventions at specific areas.

Gestures

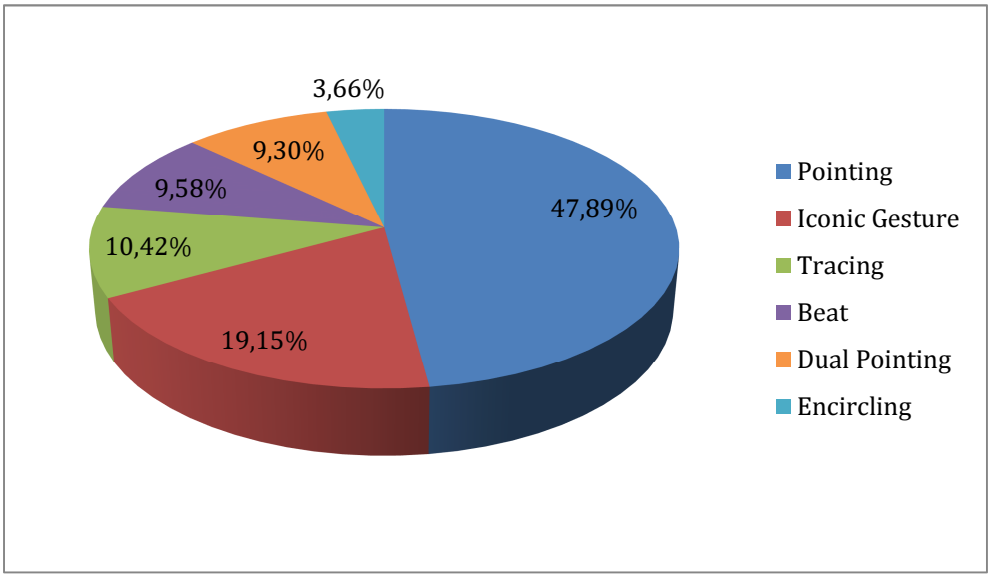


Figure 55: Distribution of gestures in the Evaluate Result of Intervention category

The participants used pointing to direct the focus of attention to the interventions they were discussing on the screen. Pointing gestures also identified the locations of virtual objects or points of interest on the map.

One focus of the evaluation process was the attributes of the virtual objects. Discussions about the shape, size, and spacing of objects were aided by iconic gestures. Tracing on the other hand highlighted the properties of flows and roads. On a few occasions dual pointing connected points on the screen with points on the map. That five of the eight dual pointing gestures originated at the projection screen and that approximately half of the pointing gestures were also directed at the screen shows that the virtual scene was the reference system for the evaluation process.

Object Manipulations

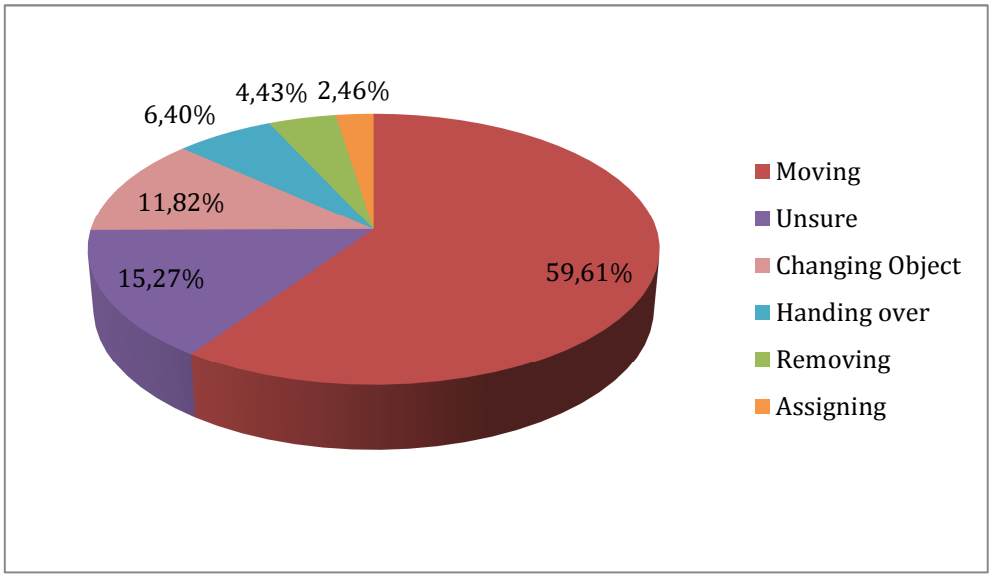


Figure 56: Distribution of object manipulations in the Evaluate Result of Intervention category

If the participants decide to modify the performed interventions based on their discussions, carrying out these changes is considered to be part of the Perform Intervention phase. Thus, the category at hand only includes 15.39% of active object manipulations. The majority of them are participants fine tuning the positioning of virtual objects, especially flows. They also make minor changes to their appearance. There is only a small amount of uncertainty reflected in their actions, which again indicates that the participants were overall very certain about their interventions.

Scene Features

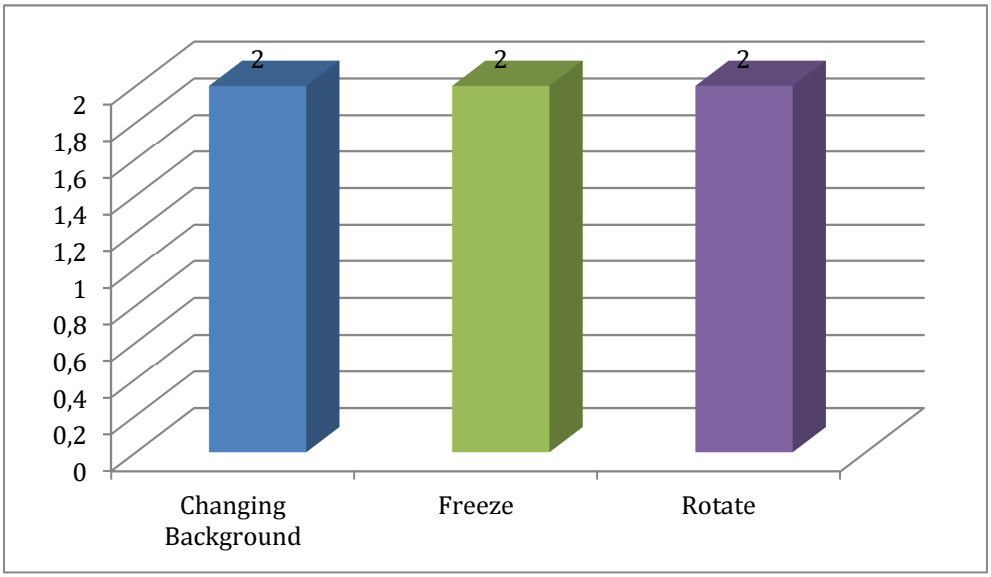


Figure 57: Usage of MR scene features in the Evaluate Result of Intervention category

The participants do not rely on different views to evaluate the scene. There are only two instances of changing the background and hence the viewpoint and rotating the panorama. It is unclear if the prepared panoramas overlap. If they do not, the same section cannot be seen from different perspectives. Having several perspectives of an area might lead the participants to consider other aspects of the urban sight, but preparing panoramas is a very time consuming process and might outweigh the possible benefits. There were two instances where, after deliberation, the participants were satisfied with their interventions and froze the scene.

9.2.5 Understand MR Scene

General

This summary is based on [1]. A mixed reality scene, as the name suggests is a combination of real and virtual elements. Even though a vast number of factors are taken into account, matters of occlusion, alignment, and positioning of virtual objects in a real scene are extremely challenging and it is nearly impossible to generate a perfect

scene where all of these matters are handled correctly. Furthermore, the process of creating a MR scene was designed to be quick and easy. The participants should be able to express their vision on the fly. The virtual content created for each workshop is only a representation of ideas and the background panoramas are an amalgamation of photographs. These elements are approximations. The scenes that result from combining them can be difficult to comprehend and are open to interpretation. For example, 2D billboards seem to be floating above the ground, because of different perspectives of the billboard and the panorama. Likewise, the size of objects could be distorted because critical 3D information was missing.

Incorporating a *depth map* into the panorama provided the 3D information necessary to handle occlusion correctly. Real objects in the foreground that are part of the panorama now obstruct virtual objects behind them. Flows also provide important visual cues. The animated pedestrians and cars show the surface of the terrain and add depth information to the scene. They are also reference points for deducing the size of an object.

The human brain also depends on shading, color, and shadows to fully comprehend a 3D space. These visual cues are not provided and throughout working with the ColorTable the participants need to actively establish an understanding of the 3D scenario. This also includes orienting themselves in the scene and searching for their interventions. The participants dealt very well with this lack of visual cues.

Changing the background always necessitates that the participants examine the scene to orient themselves. Besides being triggered by a new view, understanding activities are also interspersed in the other categories and those can be difficult to detect. Even though measures have been taken to provide visual cues, understanding the MR scene still takes up 21.43% of the time. This is almost equal to the time spent on performing an intervention. Adding further visual information might be a good idea to streamline the process of working with the ColorTable. Then again, during this phase the participants' engagement with the scene is very intense. This opens up new ways of looking at an intervention and fosters new ideas.

Use of Space

As during the Perform Intervention and Evaluate Result of Intervention phases the participants are gathered around the ColorTable in a half circle, but the formation is slightly more open and relaxed, because they move away from the table a little bit to have a better overview of the projection screen and the table. On occasion, they move closer to the table to point out a feature on the map.

Focus of Attention

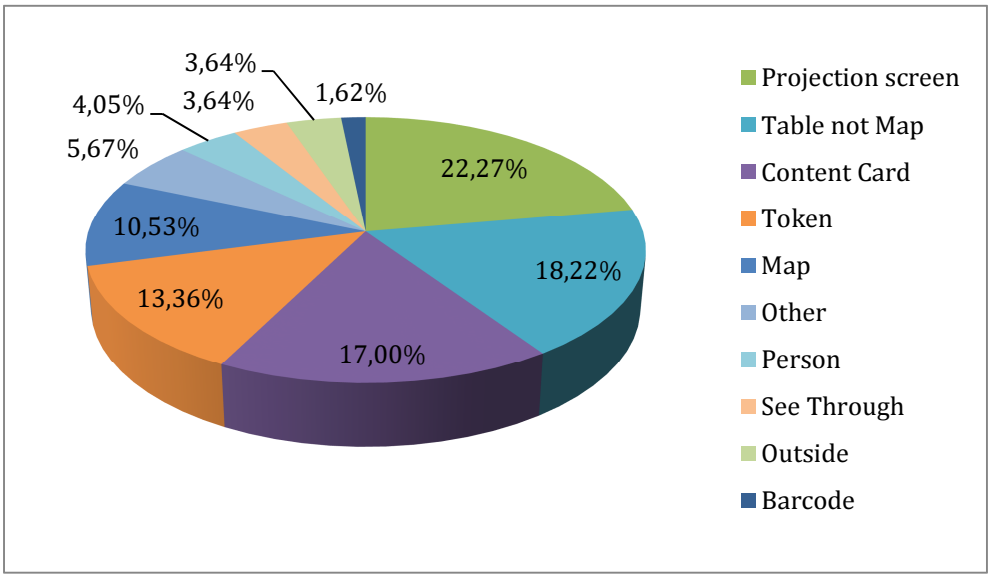


Figure 58: Distribution of focus in the Understand MR Scene category

Nearly all the components of the ColorTable play a part in comprehending the virtual scene. It must be noted that the *Table not Map* area also includes the configuration area, where the content cards are placed to assign them to tokens. Due to the language barrier it is very hard to differentiate between pointing gestures that are directed at the configuration area and those that are aimed at the content cards on it. During the Understand MR Scene phase it seems rather unlikely that the participants were pointing at a general area that in itself does not provide any information about the virtual scene. Thus, the pointing at the *Table not Map* area was very likely directed at content cards.

This means that content cards were the most pointed at features of the table, which in turn adds further credit to the observation that participants associate the virtual objects more with the content cards than with the corresponding tokens. The participants use the virtual objects as markers to guide them within the MR scenario. The pointing gestures are used to identify a virtual object that would then be located in the scene. The 13.36% of pointing gestures with a focus on tokens also fulfill this function.

Since the participants are looking for specific objects the pointing gestures at the screen are less than one might expect in this category. Only 10.53% are focused on the map, which suggests that picking precise points on the map to than look for on the screen is impractical and unnecessary to understand the MR scene.

Gestures

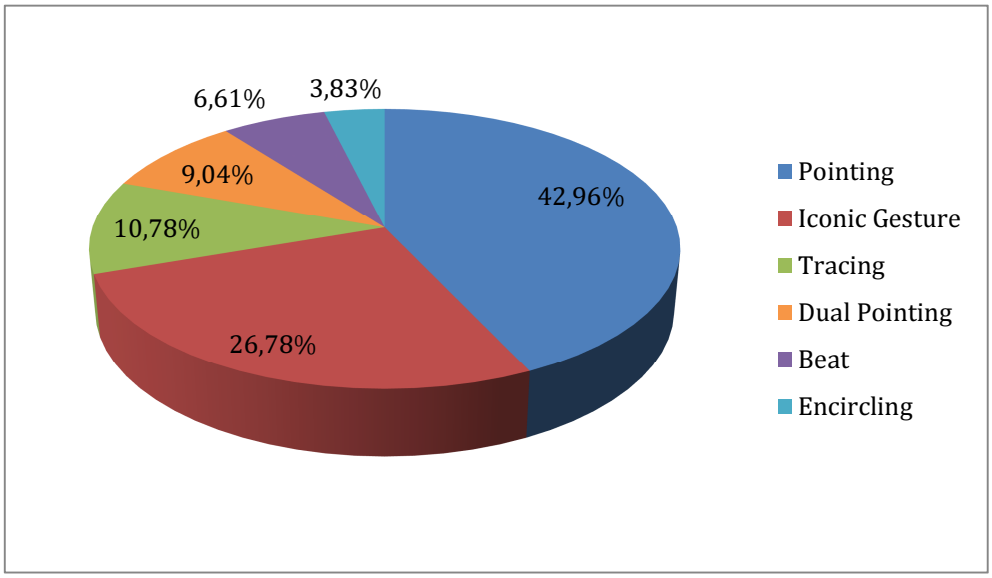


Figure 59: Distribution of gestures in the Understand MR Scene category

In the Understand MR Scene category the participants use pointing to establish common reference objects or points that help them to grasp the virtual scene. The iconic gestures describe visual features of the reference objects or the MR scenario to aid with identifying them. Furthermore, they illustrate spatial relations between

objects. The course of roads and flows includes important information about the virtual scene, tracing them brings that information into the foreground.

One would expect the number of dual pointing gestures to be more prominent during the Understand MR Scene category, because they could create direct links between points on the screen and their counterparts on the map. Dual pointing gestures are in general very rare. The entire video material only included 32 instances of dual pointing.

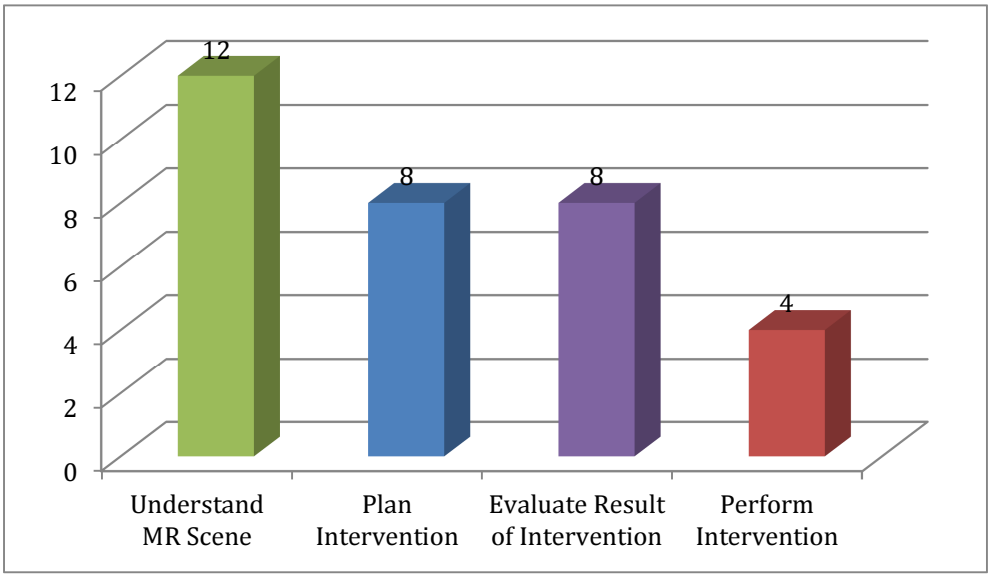


Figure 60: Numbers of dual pointing in each activity category

Even though 12 of them belonged to the Understand MR Scene category, these numbers are rather insignificant compared to the grand total of 1217 gestures. On some occasions the participants used dual pointing to establish links between the map and other points in the real world. In general, the virtual objects, which had a direct impact on the scene, established connections between the participants and the virtual realm. This is also supported by the observations above.

Object manipulations

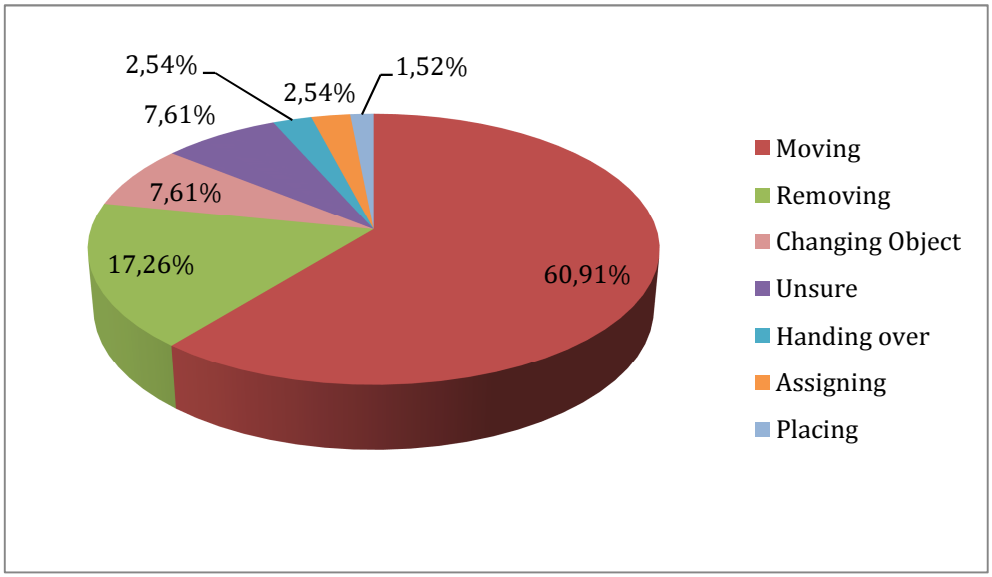


Figure 61: Distribution of object manipulations in the Understand MR Scene category

Moving objects is the predominant activity in the Understand MR Scene category, but to a certain degree the participants repurposed it. The objects are slightly repositioned to aid with finding them in the scene. The corresponding movement makes them easier to spot in the panorama or they might have even been concealed behind another object. The participants also infer from it which object is assigned to which token. Another way to find an object is to increase its size or change its color. Objects that obstruct the view might have been removed, but *Removing* predominantly includes instances were tokens and content cards were removed after a scene was frozen. There were also a few occasions when the participants had to move tokens to facilitate tracking.

Locating interventions is paramount to working with the ColorTable and enables the participants to orientate themselves in the panorama. This is additional evidence that the virtual objects are strong links to the virtual scene.

Scene Features

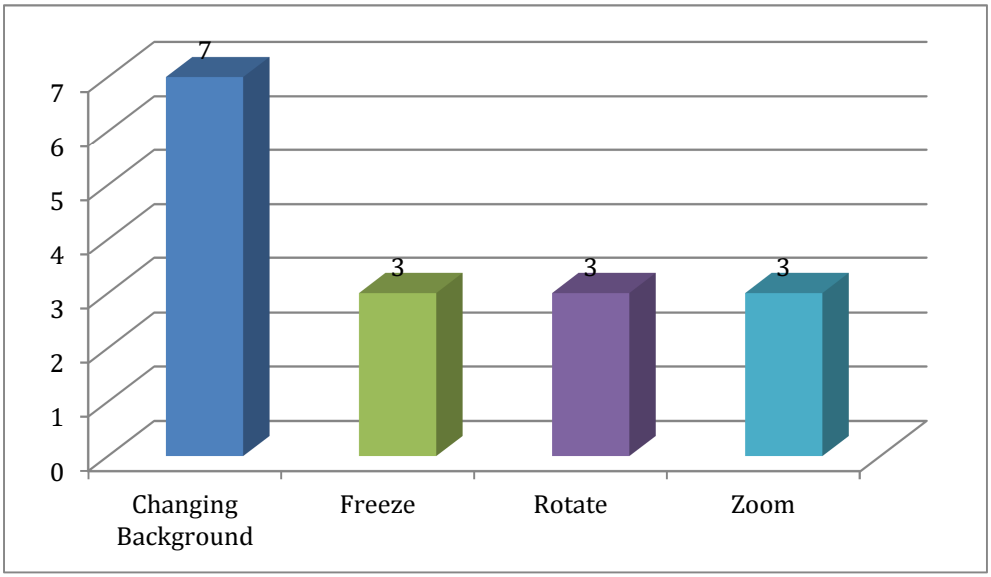


Figure 62: Usage of MR scene features in the Understand MR Scene category

Changing the background requires that the participants orientate themselves in the new scene. They would select different viewpoints to find a virtual object they had placed or to explore and comprehend the virtual area, where they wanted to perform an intervention. Rotating and zooming would allow them to navigate through the scene and get a better overview or to home in on a particular section. On some occasions the participants would freeze the scene after they had fully grasped it and all the elements were at the correct locations.

10 Conclusion

10.1 Summary

The ColorTable brings together a diverse group of stakeholders and gives them the tools to create urban planning concepts in collaboration. The idea for the ColorTable was developed as part of the IPCity project and for the final report the research team performed a qualitative analysis of working with the ColorTable. This analysis was the steppingstone for the quantitative analysis presented in this thesis, which is a close examination of the role of gestures. To start, the components and the functions of the ColorTable were described in Chapter 2.

The theoretical foundation for analyzing gestures was laid in Chapters 3, 4, and 5. They included a detailed definition, a description of how to separate them from the "background noise" of other nonverbal behavior and the visual criteria to allot them to a specific category. Because of their significance special attention was awarded to pointing gestures in Chapter 5.

A participatory workshop was the source for the data collection and the course of events was described in Chapter 6. Video observation was used as the main form of documentation among others. Chapter 7 looks at the pros and cons of using digital video.

The framework for the qualitative analysis was constructed out of eight different research questions and four distinct activity categories. The methods and relevant results of the qualitative analysis were outlined in Chapter 8.

The quantitative analysis was based on a theoretical framework that was derived from the procedures and findings of the qualitative analysis. Besides gestures, the use of space and application functions is also taken into account, resulting in the discovery of distinct behavioral patterns for each of the four activity categories in Chapter 9.

The thesis concludes with the following summary of research results and possible directions for future work.

10.2 Conclusions

The quantitative analysis confirmed the findings of the qualitative analysis. The participants depend on gestures to achieve an understanding of the site and to plan and coordinate their interventions. Pointing gestures are crucial for mapping events in the real world and the virtual realm. The quantitative analysis also garnered a few additional insights of its own.

The interventions were planned and discussed outside of the virtual realm. The participants spent most of the time in the Plan Intervention category and developed their ideas using the physical map as a reference point to the urban planning site. Only once they were certain about a decision did they translate it into the virtual scene. This course of action is also reflected in the comparatively small amounts of time spent in the Evaluate Result of Intervention and the Perform Intervention category.

That performing an intervention was not very time-consuming also shows that the participants did not have any problems with using the ColorTable. Their attention was equally divided between the three areas involved in implementing an intervention. Thus, they moved smoothly through the workflow and hardly ever hesitated or needed assistance.

In addition to representing their ideas, the virtual objects also functioned as links to the virtual scene via their physical representations. The participants moved the tokens to

find the corresponding object in the scene and use it as a marker to guide them. The content cards served both as reminders and physical handles to ideas.

Even though the quantitative analysis produced useful and interesting results, it could not have done so without being based on the qualitative analysis. The background information and findings of the qualitative analysis were an integral part, because they grounded the quantitative analysis in the context of the larger tasks, the setting of the ColorTable, and the surrounding MR-Tent. Furthermore, a laborious quantitative analysis might not have been necessary to reach the insights listed above; because they derive from aspects the qualitative analysis did not really deal with. The qualitative analysis might have come to the same insights with additional research questions.

10.3 Future Work

Pointing has a vital function when working with the ColorTable. Hence incorporating deictic gestures into the system will enhance and simplify the workflow. One possibility would be to enable participants to highlight virtual objects in the MR scene by pointing at them and using motion detection. A more general solution might be to move a "spotlight" over the scene to draw the attention to a specific point. But this "virtual pointing" should not become too powerful, for example by moving the objects via pointing. This might take away from the haptic experience of the physical map and the tokens and thus diminish the connection to the site that is created by the tangible interface.

The participants spent a lot of time discussing interventions outside of the virtual realm. The design process might benefit from different possibilities being implemented in MR scenes and thereby opening up new ways of seeing the urban site and inspiring new ideas. The Urban Sketcher and the Paper Sketching are a step into that direction, because participants can create their own custom objects. The danger is that too much time might be spent by the participants on making virtual content and that they become too focused on small details. Research should be done into whether the design

process would improve by being shifted into the virtual realm and if so, how this can be achieved efficiently.

Additional findings could be garnered by taking a closer look at iconic gestures. Examining the meaning they impart and how they are used metaphorically might uncover the thought processes that influence the urban planning with the ColorTable. The analysis at hand has shown that gestures are pointing out important insights and there is still more to discover.

Table of Figures

Bibliography

[1] V. Maquil, The ColorTable: an interdisciplinary design process, Institute of Design & Assessment of Technology, TU Vienna, 2010.

[2] T. Gevers, J. Van de Weijer and H. Stokman, "Color feature detection," in *Color Image Processing: Methods and Applications*, R. Lukac and K. N. Plataniotis, Eds., CRC Press, 2006.

[3] M. Sareika and D. Schmalstieg, "Urban sketcher: Mixed reality on site for urban planning and architecture," in *ISMAR '07: Proceedings of the 2007 6th UEE and ACM International Symposium on Mixed and Augmented Reality*, Washington, 2007.

[4] T. Binder, G. De Michelis, M. Gervautz, G. Jacucci, K. Matkovic, T. Psik and I. Wagner, "Supporting configurability in a mixed-media environment for design students," *Pers Ubiquit Comput,* pp. 310-325, 29 07 2004.

[5] V. Maquil, I. Wagner, M. Basile, L. Ehrenstrasser, M. Idziorek, B. Ozdirlik, M. Sareika, J.-J. Terrin and M. Wagner , "D6.4. Final Prototype of Urban Renewal Applications," TU Vienna, 2010.

[6] I. Wagner, M. Basile, L. Ehrenstrasser, V. Maquil, J.-J. Terrin and M. Wagner, "Supporting community engagement in the city: urban planning in the MR-Tent," in *Proceedings of Communities and Technologies, June 25 - 27*, University Park, 2009.

[7] A. Kendon, Gesture - Visible Action as Utterance, Cambridge: University Press,

2004.

[8] E. Goffman, Behavior in Public Places: Notes on the Social Organization of Gatherings, New York: The Free Press, 1963.

[9] "Noël Mamère sanctionné pour un bras d'honneur," 16 12 2009. [Online]. Available: http://www.lepoint.fr/actualites-politique/2009-12-16/affaire-greenpeace-noel-mamere-sanctionne-pour-un-bras-d-onneur/917/0/405475. [Accessed 22 09 2011].

[10] "Van Bommel wie Effenberg," 07 03 2007. [Online]. Available: http://www.n-tv.de/sport/Van-Bommel-wie-Effenberg-article216795.html. [Accessed 22 09 2011].

[11] "Mark van Bommel: Seine größten Ausraster," Zeitungsverlag tz München GmbH & Co.KG, 08 08 2009. [Online]. Available: http://www.tz-online.de/sport/fc-bayern/fotostrecke-van-bommel-ausraster-443439.html. [Accessed 22 09 2011].

[12] "Bras d'honneur," [Online]. Available: http://en.wikipedia.org/wiki/Bras_d'honneur. [Accessed 22 09 2011].

[13] D. Morris, P. Collett, P. Marsh and M. O'Shaugnessy, Gestures: Their Origins and Distribution, London: Jonathan Cape, 1979.

[14] "List of gestures," [Online]. Available: http://en.wikipedia.org/wiki/List_of_gestures. [Accessed 22 09 2011].

[15] "Gestures," [Online]. Available: http://www.statemaster.com/encyclopedia/Gestures#OK.2A. [Accessed 22 09 2011].

[16] E. Goffman, Frame Analysis, Cambridge, MA: Harvard University Press, 1974.

[17] P. Ekman and W. V. Friesen, "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding," *Semiotica,* vol. I, no. (I), pp. 49-98, 1969.

[18] D. McNeill, Hand and Mind: What Gestures Reveal about Thought, Chicago: University of Chicago Press, 1992.

[19] A. Kendon, "An Agenda for Gesture Studies," *Semiotic Review of Books,* pp. 8-12, 1996.

[20] D. McNeill, Gesture and Thought, Chicago: The University of Chicago Press, 2005.

[21] S. Kita, *The Temporal Relationship between Gesture and Speech: A Study of Japanese-English Bilinguals,* Chicago: University of Chicago, 1990.

[22] R. Mayberry and J. Jaques, *Gesture production during stuttered speech: Insights into the nature of gesture-speech integration,* D. McNeill, Ed., Cambridge: Cambridge University Press, 2000, pp. 199-214.

[23] "Stuttering," [Online]. Available: http://en.wikipedia.org/wiki/Stuttering. [Accessed 22 09 2011].

[24] J. M. Iverson and S. Goldin-Meadow, "What's communication got to do with it? Gesture in congenitally blind children," *Developmental Psychology,* vol. 33, no. 3, pp. 453-467, 1997.

[25] D. McNeill, J. Cassell and K.-E. McCullough, "Communicative Effects of Speech-Mismatched Gestures," *Research on Language and Social Interaction,* vol. 27, no. 3, pp. 223-237, July 1994.

[26] S. D. Kelly, D. J. Barr, R. B. Church and K. Lynch, "Offering a Hand to Pragmatic Understanding: The Role of Speech and Gesture in Comprehension and Memory," *Journal of Memory and Language,* vol. 40, no. 4, pp. 577-592, 1999.

[27] D. Efron, Gesture and Environment, New York: King's Crown Press, 1941.

[28] H. Brookes, "What gestures do: Some communicative functions of quotable gestures in conversations among Black urban South Africans," *Journal of Pragmatics,* vol. 37, no. 12, pp. 2044-2085, 2005.

[29] "ASLU - American Sign Language resource site," 1997 - 2009. [Online]. Available: http://www.lifeprint.com/. [Accessed 22 09 2011].

[30] "Sign of the horns," [Online]. Available: http://en.wikipedia.org/wiki/Sign_of_the_horns. [Accessed 22 09 2011].

[31] W. Wong, "Language Trainers Blog," 24 09 2007. [Online]. Available: http://www.languagetrainers.co.uk/blog/2007/09/24/top-10-hand-gestures/. [Accessed 26 07 2010].

[32] D. McNeill and E. Levy, "Conceptual Representations in Language Activity and Gesture," in *Speech, Place, and Action*, R. J. Jarvella and W. Klein, Eds., Chichester, John Wiley & Sons, 1982, pp. 271-296.

[33] A. Kendon, "How gestures can become like words," in *Cross-Cultural Perspectives in Nonverbal Communication*, Toronto, Hogrefe, 1988, pp. 131-141.

[34] G. Calbris, The Semiotics of French Gesture, Bloomington: Indiana University Press, 1990.

[35] D. Wilkins, "Why pointing with the index finger is not a universal (in sociocultural and semiotic terms).," in *Pointing: Where Language, Culture and Cognition Meet*, S. Kita, Ed., Hillsdale, New Jersey, Lawrence Erlbaum, 2003.

[36] C. Goodwin, "Pointing as Situated Practice," in *Pointing: Where Language, Culture and Cognition Meet*, S. Kita, Ed., Mahwah, NJ, Lawrence Erlbaum, 2003, pp. 217-241.

[37] "Karl Bühler," [Online]. Available: http://de.wikipedia.org/wiki/Karl_Bühler. [Accessed 22 09 2011].

[38] J. Hindmarsh and C. Heath, "Embodied reference: A study of deixis in workplace interaction," *Journal of Pragmatics,* vol. 32, no. 12, pp. 1855-1878, 2000.

[39] C. Goodwin, "Environemntally Coupled Gestures," in *Gesture and the Dynamic Dimension of Language*, S. D. Duncan, J. Cassell and E. T. Levy, Eds., Amsterdam/Philadelphia, John Benjamins Publishing Company, 2007, pp. 195-212.

[40] K. M. Murphy, "Building Meaning in Interaction: Rethinking Gesture Classifications," in *Crossroads of Language, Interaction, and Culture*, vol. 5, Regents of the University of California, 2003, pp. 29-47.

[41] C. Pearce, M. Arnold, C. Philips and K. Dwan, "Methodological considerations of digital video observation: Beyond conversation analysis," *International Journal of Multiple Research Approaches,* vol. 4, no. 2, pp. 90-99, August 2010.

[42] "www.noldus.com," Noldus Information Technology, 2011. [Online]. Available: http://www.noldus.com/human-behavior-research/products/2/the-observer-xt. [Accessed 19 09 2011].

[43] "www.studiocodegroup.com," Studiocode Business Group, 2011. [Online]. Available: http://www.studiocodegroup.com/?page_id=77. [Accessed 19 09 2011].

[44] K. Kopenhaver Haidet, J. Tate, D. Divirgilio-Thomas, A. Kolanowski and M. B. Happ, "Methods to Improve Reliability of Video Recorded Behavioral Data," *Research in Nursing & Health,* vol. 32, no. 4, pp. 465-474, August 2009.

[45] T. Coleman, "Using video-recorded consultations for research in primary care:

advantages and limitations," *Family Practice,* vol. 17, no. 5, pp. 71-77, 2000.

[46] "Spoonerism," [Online]. Available: http://en.wikipedia.org/wiki/Spoonerism. [Accessed 22 09 2011].