Présentée pour obtenir le grade de

Docteur de l'Université Paris Sud

Discipline : **Informatique**

Préparée au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI-CNRS)

Dans le cadre de l'Ecole Doctorale d'Informatique de l'Université Paris Sud (ED427)

Présentée par :

Ning TAN

Le 31 janvier 2012

Titre :

Posture and Space in Virtual Characters: Application to Ambient Interaction and Affective Interaction

Directeurs de thèse

Jean-Claude MARTIN, Professeur à l'Université Paris-Sud

Yacine BELLIK, Maître de Conférences à l'Université Paris-Sud

Rapporteurs de thèse

Elisabeth André, Professeur à l'Université Augsbourg

Pierre de Loor, Professeur à l'Ecole Nationale d'Ingénieurs de Brest

Membres de Jury

Anne Vilnat, Professeur à l'Université Paris-Sud

Stéphanie Buisine, Ingénieur de Recherche à Arts et Métiers ParisTech

TABLE OF CONTENTS

Table of contents 2
Abstract 10
I. Introduction 11
I.1 Related work and limitations of current systems 11
I.2 Thesis scope and objectives 16
I.3 Approach 17
I.4 Collaboration projects 18
I.5 Outline of the thesis 20
II. Coding and analyzing static postures of humans 21
II.1 Related work 21
II.2 Thesis contribution: the EXPO-sitting scheme for coding whole body static postures during seated dyadic conversations 33
II.3 Thesis contribution: Estimating postural convergence during dyadic conversations 45
II.4 Thesis contribution: the EXPO-standing scheme for coding standing leg postures . 62
II.5 Conclusion 75
III. Bodily expressions in virtual characters: Application to Affective Interaction 77
III.1 Related Work 77
III.2 Thesis Contribution: Designing static postural expressions of action tendencies 92
III.3 Thesis Contribution: Evaluating the static postural expressions of action tendencies 98

III.4 Thesis Contribution: Designing dynamic postural expressions of action tendencies 105

III.5 Thesis contribution: Evaluating the dynamic postural expressions of action tendencies 109

III.6 Conclusions 123

IV. Bodily Expressions in Virtual Characters: Application to Ambient Interaction 124

IV.1 Related work 124

IV.2 Designing and evaluating a location-aware virtual character in Ambient Interaction 129

IV.3 Conclusions 151

V. Conclusion and Future Directions 152

V.1 Research contributions 152

V.2 Practical contributions 154

V.3 Future directions 156

Appendices 158

References 188

LIST OF FIGURES

Figure 1: Methodology for modeling human conversation and building embodied conversational agents (Cassell 2007) 17

Figure 2: A frame from one video of the CID corpus (Bertrand, Blache et al. 2008) 19

Figure 3: Annotation Scheme for Conversational Gesture (Kipp, Neff et al. 2006) 23

Figure 4: Posture Scoring System (Bull 1987) 25

Figure 5: Identical postures (front) and mirror-image postures (back) (Bull 1987) 28

Figure 6: Gestural convergence observed in dyadic conversation (Kimbara 2002; McNeill 2005) 30

Figure 7: Subjects are wearing Optitrack markers fixed to torsos, arms and hands (hands, wrists, elbows, shoulders, sternum, and three markers mounted on a head-worn tiara (Edlund et al. 2010) 32

Figure 8: Segmentation of postures according to the Posture Scoring System (Bull 1987) 33

Figure 9. The EXPO-sitting posture coding scheme 34

Figure 10: Dentogram of the clusters based on the 89 extracted posture instances (each color represent a different posture type) 37

Figure 11: Clustering of posture types 38

Figure 12: Illustrations of three typical leg postures occurring during the first 15 minutes of conversation for subject CM. 38

Figure 13: Percentage of explanations by the principal components 39

Figure 14: Gesture space (McNeill 1992) 42

Figure 15: Cohen's kappa 43

Figure 16: An example of postural convergence for a pair of interlocutors in CID (See text for explanations) 45

Figure 17: Spatial arrangement of subject AB and subject CM and the space they share during a conversation 46

Figure 18: Annotations of the speaker / listener role using Anvil (Kipp 03) 47

Figure 19 : Principal arm configurations according to speaker / listener roles 51

Figure 20 : Main differences between speakers and listeners in terms of leg postures 54

Figure 21: Number of postural convergence for AB when she is listening, and CM when she is listening, during a one-hour dyadic conversation 58

Figure 22: Number of convergences observed according to left/right and lower/upper body 59

Figure 23: Number of postural convergence in 4 videos of the CID corpus (for a total of one hour) as a function of left/right body 59

Figure 24: Frames of the ContAct corpus 67

Figure 25: An example of leg posture annotations in the story-telling corpus ContAct 68

Figure 26: Dentogram of the clusters based on the 140 extracted posture instances from the annotations of 5 clips (different colors represent different posture types) 70

Figure 27: Mapping the 140 extracted posture instances on the 2-dimensional principal component space 72

Figure 28: A continuum proposed for representing the theories of emotions (Gross and Barrett 2011) 78

Figure 29: Models of emotion represented according to the components of emotions (lines) and the phases of cognitive evaluation (columns)(Scherer 2010). 78

Figure 30: Computational models of emotions and their relations with the underlying psychological theories (left) (Marsella, Gratch et al. 2010). 79

Figure 31: Frames from the PERMUTATION corpus. 92

Figure 32: Bodily expressions designed for the five selected action tendencies and for the Neutral.....95

Figure 33: An example of the questionnaire for the recognition of action tendencies 99

Figure 34: An example of the questionnaire for the attribution of emotion categories 100

Figure 35: Relations between the four action tendencies (IN: in command, AG: Antagonistic, EX: Exuberant, DFV: Disappear from View) on the scales of intensity of movement and temporal extent.....106

Figure 36: Combining opposing force, intensity of movement and temporal extent for specifying the dynamics of our postural expressions of action tendencies 107

Figure 37: Effect of Segmentation and Duration of the Animation within the action tendency Disappear from View (comparison of mean scores of perceived action tendencies). 116

Figure 38: Effect of Segmentation and Duration of the Animation within the action tendency Antagonistic (comparison of mean scores of perceived action tendencies). 116

Figure 39: Effect of Segmentation and Duration of the Animation within the action tendency in command (comparison of mean scores of perceived action tendencies). 117

Figure 40: The ambient intelligent room used in the experiment with Ubisense location sensors at the four corners of the room (left); a MARC virtual character is displayed on the TV display (right) 130

Figure 41: Some illustrations of postures designed and validated for the following communicative acts (from left to right): point, disapprove, congratulate.....132

Figure 42: The automata managing the interaction between the user, the objects, the ambient system and the virtual character: a) interacting with the non-adaptive virtual character b) interacting with the adaptive virtual character. 133

Figure 43: The ambient intelligent room with the virtual character displayed on a TV set, a user equipped with a Ubisense tag, a target object on the table, and a location sensor on the ceiling.....135

Figure 44: Annotation of the behaviors of a user with the Anvil tool (Kipp, 2003) 137

Figure 45: Structural model explaining relationships between the conditions of virtual character, perceived social presence, user characteristics and user engagement.....146

Figure 46 : Moderation by Gender.....147

Figure 47 : Moderation by virtual character experiences 148

Figure 48 : Moderation by Ambient Experiences.....148

Figure 49: Future directions of the work 157

Figure 50: Effect of Segmentation and Duration of the Animation within the action tendency Disappear from View (comparison of mean scores of perceived action tendencies). 162

Figure 51: Effect of Segmentation and Duration of the Animation within the action tendency Antagonistic (comparison of mean scores of perceived action tendencies). 162

Figure 52: Effect of Segmentation and Duration of the Animation within the action tendency in command (comparison of mean scores of perceived action tendencies). 163

Figure 53: Virtual characters are used as ambient interaction metaphor in iSpace 182

Figure 54: Steps for the role-playing 183

Figure 55: Given the same communicative act (inform cooking time), the agent exhibit different poses (from left to right): 1) the recorded subject; 2) the extravert agent; and 3) the introvert agent. 187

LIST OF TABLES

Table 1: Outline of the thesis 20

Figure 2: Body motion illustrations (Bartlett 1997) 26

Table 3: Default postures in different disciplines: a) fundamental, b) anatomical, c) ergonomic 27

Table 4: Postures annotated for different body parts in the CID corpus (video AB_CM): number of annotated segments and their total duration in seconds 36

Table 5: PCA scores: transforming the original data of leg posture instances into the space of the first 6 principal components 40

Table 6: Average kappa values for each categorical item of GestureSpace for three coders 43

Table 7: Duration of annotated postures according to the conversation role: in terms of number of frames and seconds (25 frames per second) 47

Table 8: Measures of association between the speaker / listener role and the attributes of the upper body 48

Table 9: Arm radial orientation 49

Table 10: Arm swivel 49

Table 11: Arm touching 49

Table 12: Arm height 49

Table 13: Arm distance 50

Table 14: Association between the speaker / listener role and the attributes of the lower body (significative results in bold) 52

Table 15: Leg-to-leg distance 52

Table 16: Leg-crossing forms 53

Table 17: Leg orientation 53

Table 18: Leg swivel 53

Table 19: Kendon's Alignment of Gestural and Intonational Hierarchies (Loehr 2004) 55

Table 20: Numbers of postural convergence in 4 videos of the CID corpus (for a total of one hour) 57

Table 21: Main attributes where postural convergence is observed 60

Table 22: The EXPO-standing coding scheme 62

Table 23: Illustration of values for the leg-role attribute for the left leg 63

Table 24: Illustration of values for the leg-position attribute for the right leg 63

Table 25: Illustration of values for the leg-shape attribute for the right leg 64

Table 26: Illustration of values for the knee-shape attribute for the left leg 64

Table 27: Illustration of values for the foot-orientation attribute for the right leg 65

Table 28: Illustration of values for the foot-orientation attribute for the left leg 65

Table 29: An example of annotation of leg postures using the EXPO-standing coding scheme 66

Table 30: A second example of annotation of leg postures using the EXPO-standing coding scheme.....66

Table 31: Corpus descriptions 67

Table 32: A summary of leg posture annotations based on the 5 recording clips 68

Table 33: Kappa values based on the annotations of leg postures by two coders 69

Table 34: Number of instances of leg postures before and after the extraction 70

Table 35: Percentage of the three posture types (posture type #1, posture type #2 and posture type #3) expressed by subjects (S2, S3, and S4).....71

Table 36: Descriptions of the 3 posture types based on the 140 extracted posture instances..73

Table 37: PCA scores: transforming the original data into the space of the principal components (the values in bold are those of highest absolute value per component). 74

Table 38: Prediction of five emotions with the best viewpoint and the corresponding anatomical features (Coulson 2004) 81

Table 39: Predictions of body changes following major SEC outcomes (Scherer 2010).....82

Table 40: Laban Movement Analysis (Body, Effort, Shape and Space) 84

Table 41: Studies and corpora of bodily expressions of emotions (1).....87

Table 42: Studies and corpora of bodily expressions of emotions (2).....88

Table 43: A review of expressivity parameters in motor behavior and computational generation systems 91

Table 44: The five action tendencies selected for our study and the written description proposed by Frijda (Frijda, Kuipers et al. 1989).....93

Table 45: Annotations of bodily expressions of action tendencies (for each body part, the most frequent annotations are listed) 94

Table 46: Specification of some of the postures using the EXPO scheme 96

Table 47. Attribution rates for target and distracting postures for each action tendency. 101

Table 48. Attribution of emotion categories to the postural expressions of action tendencies. The emotions with a level of attribution of more than 10% for each posture are in bold. The percentages for the emotion categories that are predicted by Frijda 1989 for each action tendency are displayed in grey shading. 103

Table 49: Mapping between action tendencies, intensity of movement and temporal extent106

Table 50: Timing parameters for the animations of the action tendencies 108

Table 51 : Factorial design of the 2nd experimental study 111

Table 52: Validation of the effects of independent factor Movement Quality on the different dependant factors 112

Table 53: Validation of the effects of independent factor Segmentation and Duration of the Animation on the different dependant factors 112

Table 54: Pairwise comparison of 5 action tendencies: the estimated difference in means and a confidence interval of the difference 114

Table 55 : Attribution of emotion categories to the dynamic bodily expressions of action tendencies (the emotions with the highest level of attribution for each expression are in bold; the predictions by Frijda are in grey shading). 121

Table 56: The behaviors of the virtual character in the two experimental conditions: non-adaptive character vs. adaptive virtual character. 132

Table 57: Collected data. 136

Table 58: The coding scheme defined for guiding the manual annotation of the collected videos of user's behaviors 136

Table 59: T-test results: comparing looking behavior of users in the two groups (the values of significance are marked in grey shading) 141

Table 60: Correlations between user's behavior and items from the perception questionnaire (the significant correlations are marked with an asterisk) (FQ: frequency, DR: duration) ... 143

Table 61: Significant correlations between independent variables and dependant variables 145

Table 62 : Significance of slopes for moderation by Gender (* is significant at p<.05) 147

Table 63 : Significance of slopes for moderation by Virtual Character Experiences (* is significant at p<.05) 148

Table 64 : Significance of slopes for moderation by Ambient Experiences (* is significant at p<.05) 148

Table 65: The main episodes and the related communicative acts 184

Table 66: Distinctive nonverbal behavior between extroversion and introversion (Gallaher 1992) 185

Table 67: General-purpose communicative functions and related expressions for the scenario "cooking pasta" (Bunt 2009) 186

Table 68: Dimensions-specific functions and related expressions for the scenario "cooking pasta" (Bunt 2009) 186

Table 69: A summary of designing bodily expressions 187

ABSTRACT

Multimodal communication is key to smooth interactions between people. However, multimodality remains limited in current human-computer interfaces. For example, posture is less explored than other modalities, such as speech and facial expressions. The postural expressions of others have a huge impact on how we situate and interpret an interaction. Devices and interfaces for representing full-body interaction are available (e.g., Kinect and full-body avatars), but systems still lack computational models relating these modalities to spatial and emotional communicative functions.

The goal of this thesis is to lay the foundation for computational models that enable better use of posture in human-computer interaction. This necessitates addressing several research questions: How can we symbolically represent postures used in interpersonal communication? How can these representations inform the design of virtual characters' postural expressions? What are the requirements of a model of postural interaction for application to interactive virtual characters? How can this model be applied in different spatial and social contexts?

In our approach, we start with the manual annotation of video corpora featuring postural expressions. We define a coding scheme for the manual annotation of posture at several levels of abstraction and for different body parts. These representations were used for analyzing the spatial and temporal relations between postures displayed by two human interlocutors during spontaneous conversations.

Next, representations were used to inform the design of postural expressions displayed by virtual characters. For studying postural expressions, we selected one promising, relevant component of emotions: the action tendency. Animations were designed featuring action tendencies in a female character. These animations were used as a social context in perception tests.

Finally, postural expressions were designed for a virtual character used in an ambient interaction system. These postural and spatial behaviors were used to help users locate real objects in an intelligent room (iRoom). The impact of these bodily expressions on the user's performance, subjective perception and behavior was evaluated in a user study

Further studies of bodily interaction are called for involving, for example, motion-capture techniques, integration with other spatial modalities such as gaze, and consideration of individual differences in bodily interaction.

Keywords: Bodily Expressions, Postures, Space, Multimodal User Interfaces, Virtual Characters, Ambient Interaction, Multimodal Corpora, Spatial Behaviors, Experimental Studies

I. INTRODUCTION

This thesis is entitled "Posture and Space in Virtual Characters: Application to Ambient Interaction and Affective Interaction". In this introduction, we start by identifying the different research areas that are relevant to this topic and summarizing the main limitations of existing systems in terms of how they consider posture and space.

We namely describe two application domains (ambient interaction and affective interaction). We explain why they are relevant for studying postures and space and discuss the research question they raise. We then describe the research goals of this thesis and the approach that we used in this thesis.

I.1 Related work and limitations of current systems

I.1.1 Multimodal Human-Computer Interaction

I.1.1.1 Human-Computer Interaction and spatial context

Human-Computer Interaction is the "process through which human users work with interactive computer systems" (Parker 2003). One major concept that needs to be considered when designing human-computer interfaces is the context in which the interaction occurs. Context is "not simply the state of a predefined environment with a fixed set of interaction resources. It is part of a process of interacting with an ever-changing environment composed of reconfigurable, migratory, distributed, and multi-scale resources" (Coutaz, Crowley et al. 2005).

A key concept that links humans to the context of interaction is space. According to Kant, space is "part of an unavoidable systematic framework for organizing our experiences". Space in interaction includes the areas defined by the orientation the movements of the participants' bodies (Rodrigues, Kopp et al.). Space is related to multiple levels of the interaction: engagement in interaction, common ground, size of the bodies, body orientations, individual differences (age, gender etc.), physical environment, and affects (Rodrigues, Kopp et al. ; Ekman and Friesen 1967; Holler 2007). Bodily experiences in the world and the mental representations of space are reciprocally linked to each other. Several studies considered different kinds of interaction space, such as personal space (Hall 1963; Wallbott 1998; Sweetser and Sizemore 2008), gesture space (McNeill 2005), and interactional space (Sweetser and Sizemore 2008).

I.1.1.2 Multimodal interfaces

In user-centered approaches to Human-Computer Interaction, the term modality refers to "a mode of communication according to human communication cues (e.g. vision, audio etc.), or according to input devices (e.g. camera, haptic sensors, and microphones)" (Bellik 1995).

Multimodal input interfaces "process combined natural input modes, such as pen, touch, speech, hand gestures, eye gaze, head and body movements in a coordinated manner with multimedia system output" (Oviatt 2002). Some multimodal input systems were observed to increase the usability of human-computer interaction, namely because they provide users with more than one available communication channel (e.g. such as speech, gesture, writing) (Oviatt 1996; Mayer 2002). Multimodal interfaces usually attempt to make input modes mapped to natural human communication modalities (Oviatt 2002).

Concerning the output side of multimodal systems, researchers investigate the presentation modes that are most suitable for different situations (Rist and Andre 1993; Rousseau 2006), different data to be presented, and how different presentation modes should be combined (globally known as "the media allocation problem"). These studies on multimodal information presentation explore how a chunk of information can be presented through several presentation modes, spatially coordinated, such as a combination of text, speech, graphics, images, etc (Hooijdonk 2008).

I.1.1.3 Virtual characters

A virtual character (Cassell, Sullivan et al. 2000) is a multimodal human-computer interface that combines gestures, facial expressions and speech to enable face-to-face communication with users. This virtual character is supposed to be endowed with conversational capacities inspired by human communication such as the ability to manage turn-taking (Cassell, Torres et al. 1998). Different terms are used depending on the underlying model and generation capacities: "Embodied Conversational Agent" (ECA), "life-like character" (Kruppa, Spassova et al. 2005), "animated agent" (Cassell, Sullivan et al. 2000), and "virtual character" (Rist and Andre 2003). Different terms underlie different meanings from canned animations to dynamic generation at run-time, and from manually controlled avatars to autonomous characters (Cassell et al., 2001).

Virtual characters might enable a flexible and natural communication with users and an optimization of the cognitive load in users because they rely on the communication modes that we use everyday for expressing various communicative functions.

There is a growing interest in HCI for body-based interfaces (Fogtmann, Fritsch et al. 2008; Castellano, Pereira et al. 2009). Several interactive designs involve the body as the user input modality (i.e. gestures, gaze, and body movements). Bodily interaction in these applications is often unidirectional. Users can send information through their body to the systems, but cannot perceive bodily signals from the systems. Several attempts are nevertheless made to display multimodal information through the body of virtual characters in ambient environments, such as a virtual dancer (Reidsma, Heylen et al. 2006), a virtual conductor (Bos, Reidsma et al. 2006), a virtual trainer (Ruttkay, Zwiers et al. 2006), a virtual anatomy assistant (Wiendl, Ulhaas et al. 2007), a virtual cooking assistant (Miyawaki and Sano 2008). For example virtual rapport agents are able to express backchannels to users using head nods (Huang, Morency et al. 2011). A few system feature virtual characters that are able to display

bodily expressions of attitudes (Ballin, Gillies et al. 2004) or proxemics (Rehm, André et al. 2005).

Yet, the use of the whole body of the virtual agent for expressing convergence, the different components of emotions or managing a shared space of interaction during a task with the user remains limited.

Research in virtual characters considers the necessary mappings between signals in multiple modalities (e.g. facial expressions, gaze, gesture, posture) and communicative functions (e.g. emotions, emphasis and etc.) (Poggi and Pelachaud 2000; Pelachaud 2005). Designing virtual characters requires informing these relations between communicative functions and their possible multimodal expressions. Although numerous studies are described in the literature, they are context specific and their results are often difficult to apply to another context. This requires additional empirical studies about nonverbal behaviors and the collection of detailed behavioral data.

I.1.2 Methodological approaches

I.1.2.1 Bodily interaction in Human Sciences

Numerous studies in the Human Sciences have explored non-verbal behaviors that occur during social interactions, for example, facial expressions (Ekman and Friesen 1975) and hand gestures (Kendon 2004). Researchers in multimodal interfaces and virtual characters often inspire from these studies. For example, Kipp (Kipp, Neff et al. 2006) developed an annotation scheme for conversational gestures that he applied to gesture generation in virtual characters.

Some studies considered full bodily expressions. Bull (Bull 1987) proposed a Posture Scoring System for describing both static and dynamic body postures. He applied it to several studies of dyadic interaction including two people sitting in front of each other. More recently, motion capture corpora were collected for studying bodily expressions of affects (Bianchi-Berthouze and Kleinsmith 2003).

The body modality is of importance in human social interactions because it supports 1) the convergence of representations of meaning across interlocutors during (Lakin, Jefferis et al. 2003; van Baaren, Holland et al. 2003); 2) the management of social space (Rodrigues, Kopp et al. ; McNeill 2005), and 3) the expression of emotions (Bianchi-Berthouze and Kleinsmith 2003; Coulson 2004; Kret, Pichon et al. 2011).

Several channels convey spatial behaviors between individuals during social interactions: pointing, body orientation, trunk leaning, avoidance / approach, etc. Researchers in social sciences categorize these nonverbal cues into several functional groups such as deictic gestures and proxemics. A deictic gesture is about "indicating an object, a location, or a direction, which is discovered by projecting a straight line from the furthest point of the body part that has been extended outward, into the space that extends beyond the person" (Kendon 2004).

Ekman considered that body actions might provide information about the intensity of the felt emotion (Ekman 1965). Other researchers such as Wallbott observed discriminative features of emotion categories in both posture and movement quality (Wallbott 1998). Nevertheless,

additional studies and multimodal data seem required to provide answers to questions such as: How do people perceive others' bodily expressions in terms of emotions? How are other components of emotion (e.g. action tendency) conveyed by the different modalities?

I.1.2.2 Multimodal Corpora

Multimodal corpora approaches use manual and/or automatic annotations of videos and other related media data (Kipp, Martin et al. 2009). Multimodal corpora can use videos that are collected in a human-human interaction context (field studies, laboratory recordings, television, movies) as well as during human-computer interactions (using a real system or a simulated system in a Wizard of Oz protocol). Studies of human nonverbal cues originally involved manual textual annotation. Recent computer-aided annotation tools, e.g. Anvil by Kipp (Kipp 2001) assist coders in the manual annotation of nonverbal behaviors that are observed in videos. These annotations are anchored in time on multiple layers. As a preliminary step to annotation, different coding schemes have to be defined at different levels of abstraction for representing multimodal behaviors. Consistency between the annotations done by different coders has to be assessed for validating the coding scheme and the annotation protocol at least on a subset of the data. Researchers also developed automated tools to capture nonverbal behaviors using image processing and motion capture that can be integrated with manual annotations.

During spontaneous conversations, multiple modalities such as speech, gesture, posture and gaze are combined in sophisticated ways (Loehr 2004). Yet, multimodal corpora studies on the way in which the different nonverbal modalities interact remain scarce, partly due to the lack of accessible and relevant resources. For example, the multimodal corpora that are currently available in French are limited in terms of number of modalities, accessibility, and spontaneity. In addition, studying relations between those modalities requires the definition of reliable coding schemes, describing multiple levels from the phonological level to the gestural and postural levels. The number of such levels of annotation at multiple levels of expression in existing corpora is also limited.

I.1.3 Applications domains

I.1.3.1 Ambient Interaction

Ambient Intelligence (Fontana, Richley et al. 2003) is characterized by three concepts: 1) ubiquitous computing: microprocessors are embedded in everyday objects that traditionally lack any computing ability (e.g. books, clothes), 2) ubiquitous communication: objects are endowed with wireless communication abilities, rely on energy sources that provide them with autonomy, and are capable of spontaneously interoperating with other objects, 3) intelligent user interfaces: human users are able to interact with these objects in a natural way (e.g. using gestures), and the objects must take users preferences and context into account.

In an ambient intelligence system, intelligent computation should be embedded in our everyday environments through a pervasive transparent infrastructure (consisting of a multitude of sensors, actuators, processors and networks) which is capable of recognizing, responding and adapting to individuals in a seamless and unobtrusive way (Ducatel, Bogdanowicz et al. 2001). The context of ambient interactions consists of the physical context (objects in the environment), the perceptual context (events or actions of the other

users, projected/expected actions of the users), the conversational context (current conversational roles of the participants) and the social context (social roles of the users, e.g. visitor in a museum, employee in a company).

The contextual information contains both static information (that does not change during the interaction) and dynamic factors (that change during the interaction) (Löckelt, Becker et al. 2002). There is a need to map the dynamic contextual information onto the relevant output modalities.

The output modalities that represent ambient information should enable to manage the shared space between users and intelligent embedded agents. Attempts that were made to address this issue were limited to spatial behaviors such as managing proximity. For example, the MirrorSpace (Roussel, Evans et al. 2004) prototypes an ambient awareness display for video communication, which uses physical proximity to control the blurring of the displayed image. The image becomes clear when a user gets close to it, and becomes blurry when he/she stands away. The concept is based on the concept of personal space (Hall 1966): the closer users stand to a device, the more engagement they desire.

These ambient systems require the definition of dedicated evaluation protocols. Users might worry about a possible loss of control of the ambient system, about the complexity of installing and maintaining these smart home systems, or about any security attacks that would grant access to their recorded private data. Researchers consider several interaction design criteria to address these issues, for example information capacity, notification level, representational fidelity and aesthetic emphasis (Pousman and Stasko 2006).

Virtual characters involve promoting social presence of a life-like entity under contexts of ambient interactions (Bailenson et al., 2001; Blascovich, 2002; McQuiggan et al., 2008). This rises additional questions: How to endow virtual characters with the ability to adapt dynamically to user's behaviors in ambient environments? How do users feel when being in the presence of a virtual character in an ambient environment? How does their perception of the spatial behaviors displayed by virtual characters (including its posture) affects their own bodily expressions?

I.1.3.2 Affective and Social Computing

Another domain that sounds relevant for studying posture and space is affective computing. Affective computing refers to the study and development of systems and devices that can "recognize, interpret, process, and simulate human affects" (Picard 1997). The machine needs to be capable of adapting and responding to the affective states of users, as this capacity is supposed to be perceived as natural, efficacious and trustworthy (Pantic and Patras 2006). One main challenge of affective computing is to foster socially situated communication with users.

Affective Computing research inspire from several approaches to emotion that are developed in Psychology. For example, the cognitive approach to emotions considers that an emotion results from a cognitive evaluation of the current context / emotional situation (Scherer 2010).

Virtual characters need to express emotions during affective interactions. However, this capacity raises several research questions: How to endow virtual characters with the ability to express complex emotions in multiple modalities? How does the user perceive these multimodal expressions? Efforts in that direction have explored how virtual characters can

express emotions via voice, facial expressions and gestures. Less is known about how their whole postures and bodily expressions might communicate emotions (Bianchi-Berthouze and Kleinsmith 2003).

In summary, current multimodal systems and virtual characters remain limited in how they consider whole bodily interaction. For example, an adequate use of bodily interactions in ambient environments should foster the consideration of spatial relationships between users and systems in a situated and social way. Similarly, the study of full body expression of emotion remains scarce when it comes to other components of emotion than categories and arousal.

Few studies in Human Sciences investigate the extent to which bodily expressions are involved in the management of the interaction space shared by two persons or convey emotions. There is a need for new corpora and analyses that enable the exploration of these bodily expressions during interaction.

Hence, two promising domains seem relevant for bodily interaction: ambient interaction and affective interaction.

I.2 Thesis scope and objectives

The goal of this thesis is to improve the interaction between virtual characters and users by exploring bodily expressions in the context of two promising application domains: ambient interaction and affective interaction.

We aim to **lay the foundations for future computational models that will enable selecting adequate bodily expressions of a virtual character to express two major and complementary communicative functions**: 1) **communicating about the management of the interaction space** shared by users and virtual characters, 2) **expressing the emotions** of the virtual character.

This goal can be broken down into three experimental research questions:

- Human-human interaction: how do bodily expressions occur in dyadic conversation?
- Virtual characters expressions: Do users perceive the emotions expressed by virtual characters?
- Users interacting with a virtual character: How do users perceive the spatial relationships with a virtual character?

We chose to focus on bodily expressions and we did not consider in details how it interacts with other modalities such as speech or facial expression. This study of multimodal signs is beyond the scope of this thesis.

In terms of data annotation, we focus on manual annotations of videos. Using motion capture techniques has been increasing the last years, but it is beyond the scope of this thesis.

I.3 **Approach**

In this thesis, we analyze the bodily behaviors with respect to the three areas of analysis:

1. humans in dyadic conversation
2. users perceiving bodily expressions of emotions
3. users in interaction with ambient intelligence systems

For each of these areas, we inspired from the classical method by (Cassell 2007) that is widely used in virtual agent research (Rehm and André 2008) (Figure 1).

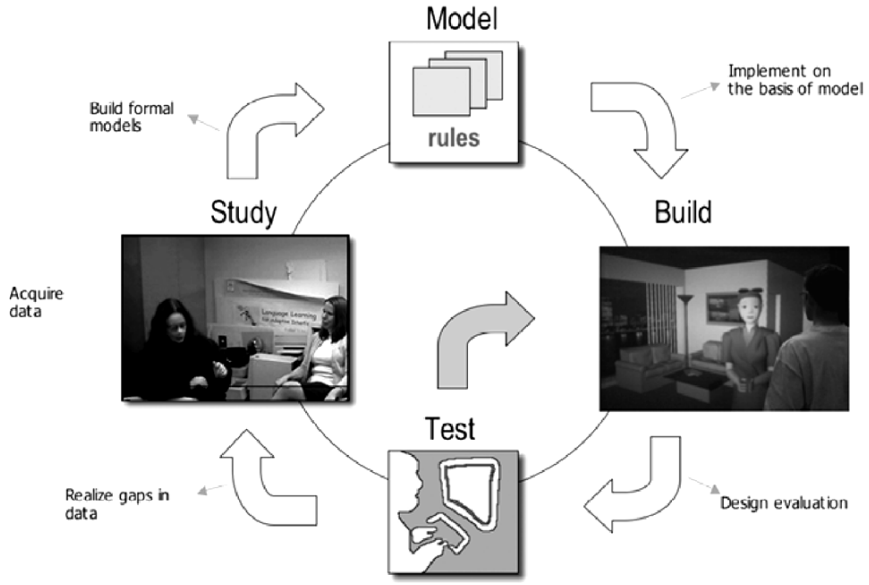


Figure 1: Methodology for modeling human conversation and building embodied conversational agents (Cassell 2007)

Multimodal Corpora. We use video corpora to study situated and bodily expressions in the three areas of analysis. In each of these areas, we annotate bodily expressions manually.
In the first area of analysis, we use a corpus of dyadic conversation in French. We proposed a coding scheme for full body postures and used it to annotate the corpus. We implemented algorithms to automatically process manual annotations of postures. We illustrate how this helps us to study an interactional phenomenon that occurs in dyadic conversation: convergence.
In the second area of analysis, we used a corpus of American television comedy-drama series to inform the design of bodily expressions of emotions.
In the third area of analysis, we recorded an interaction corpus, in which a location-aware virtual character communicates with ambient users. We used this corpus to understand how users interact with an ambient virtual character in terms of engagement.

Model. We model bodily expressions based on the literature and corpora. In the 2nd area of analysis, we proposed a model of bodily expressions that convey emotions. In

the 3rd area of analysis, we model bodily expressions that are expressed by a location-aware virtual character in interaction with users in an ambient environment.

Build. We implemented our models using the MARC virtual agent platform (Courgeon, Martin et al. 2008). In the second area of analysis, we designed a library of bodily expressions that can be used by a virtual character to express emotions. In the third area of analysis, we designed a module managing the interaction between a user and a virtual character in an ambient environment.

Test. In the second area of analysis, we analyzed how human subjects perceive the bodily expressions of emotions expressed by a virtual character. In the third area of analysis, we investigated how users interact with a location-aware ambient virtual character.

I.4 Collaboration projects

This thesis was funded by two projects. Part of the work described in this thesis was conducted in collaboration with the corresponding partners.

I.4.1 The OTIM Project

The OTIM project (Outils pour le Traitement de l'Information Multimodale / Tools for the Processing of Multimodal Information) is funded by the French Research Agency (ANR). The partners include the Speech and Language Laboratory of the University of Aix-en-Provence (LPL, coordinator), the Interactional Language Center of the University of Nantes and the French Linguistics Center of the University of Paris 3.

The aim of this project is to define a framework including conventions and tools to allow researchers to annotate multimodal information. The project builds upon a digital corpus of multimodal behaviors occurring during spontaneous conversations in French: the Corpus of Interactional Data (CID) (Bertrand et al., 2008). CID is an original audio and video database of spontaneous spoken French. Eight pairs of native French speakers take part in a one-hour dyadic spontaneous spoken conversation, in which they tell each other about some unexpected events that they experienced. Numerous annotations have been performed by the LPL and the OTIM partners at multiple linguistic levels (phonetization, syntax, prosody, discourse, hand gestures). Our work on the coding scheme for body postures was conducted within this project. In CID the files are named as follows: subject1-subject2-beginning-end (for example, the video AB-CM-0-15 was used in several annotations described in the next chapters).



Figure 2: A frame from one video of the CID corpus (Bertrand, Blache et al. 2008)

I.4.2 The ATRACO Project

Part of our work was carried out in the framework of a European Project: Adaptive and TRusted Ambient eCOlogies (ATRACO). The collaboration project partners include University of Ulm (Germany), University of Essex (UK), Computer Technology Institute (Greece) and an industrial partner inAccess Networks (Greece). This project aimed to contribute to the realization of trusted ambient ecologies. Ambient ecologies refer to space populated by connected devices and services that are interrelated with each other, the environment and the people, supporting the users' everyday activities in a meaningful way.

This thesis contributed to the Atraco project by designing a virtual character that was used in an intelligent room at LIMSI: the iRoom (Bellik, Reba ï et al. 2009). We conducted an experiment with users and evaluated the impact of the spatial behaviors of the virtual characters on the users (Tan, Pruvost et al. 2010).

I.5 Outline of the thesis

The remainder of this thesis includes three parts (Table 1).

Table 1: Outline of the thesis

			
	II. Coding and analyzing bodily behaviors of humans	**III. Bodily expressions of emotions in virtual characters**	**IV. Bodily expressions and spatial behavior in human-computer interaction**
Related work	Section II.1	Section III.1	Section IV.1
Modeling and Design	Section II.2	Section III.2	Section IV.2
Evaluation	Section II.3	Section III.3 Section III. 4	Section IV.3

Section V includes a global discussion, a general conclusion and a description of possible future directions of research.

II. CODING AND ANALYZING STATIC POSTURES OF HUMANS

II.1 Related work

Posture supports multiple important communicative functions conveying interpersonal attitudes (Argyle 1975), emotions (Wallbott 1998), communicative styles (Richmond and Croskey 1999), and personalities (Ibister and Nass 2000).

In this related work section, we do not focus on these various communicative functions, but instead we consider how researchers code and represent body postures. Chapter III will survey how postures convey emotions.

II.1.1 Conversational analysis

Conversational analysis is a sociological and socio-linguistic approach to the study of communicative patterns (Goodwin and Goodwin 2000). Conversational analysis is driven by data. Based on the detailed analysis of audiotapes and/or videotapes, researchers investigate how people, through procedural rules, engage and succeed in conversations (Sacks, Schegloff et al. 1974). Conversation involves two fundamental alternative and overlapping participative roles: speaker and listener. Conversational analysis is generally based on this basic distinction of roles in a conversation. Conversation Analysis researchers pointed out the need to study the situation in which the communication between speakers and listeners takes place. This situation includes the other participants, joint activities, the physical environment, and the interactional memory. Goodwin insists on the meaning that can be conveyed by the material structure of the environment and the surrounding space. In one of his illustrative example extracted from a case study, two girls play together. The structure of the grid drawn on the ground needs to be considered to interpret the gestural and multimodal signals sent by the girls. For example, when the girl who is listening to the other girl changes her body orientation and looks down, the other girl who was speaking using hand gesture switches to other modalities and points to the grid using her foot.

II.1.2 Gesture studies

Among the various bodily channels, hand gestures have been the focus of attention of several researchers. Gesture refers to “a movement of the body, or any part of it, that is expressive of thought and feeling” (McNeill 2005). McNeill also considers that gestures are the movements of the arms and the hands that are closely synchronized with the flow of speech (McNeill 1992). Kendon (Kendon 1980) defines gesture as a visible bodily action that plays a role in

utterances. He considers gesture as an action that has the features of manifest deliberate expressiveness. Both researchers focus on hand and arm gestures.

McNeill defined a methodology for coding the different aspects of gesture: hand form (handedness, shape of the hand, palm and finger orientation, place in gesture space), hand motion (shape of the motion, place in space, direction), meaning of the hand, and meaning of the motion (McNeill 2005). The methodology for annotating gestures features multiple steps such as: the segmentation of a gesture unit into different gestures, and the use of a pre-analysis of video data to define a gesture lexicon within a categorical transcription approach, for example, (cf. the gesture coding manual by S. Duncan[1] and gesture lexicon approaches Calbris 2003, (Kipp 2004)). Lexicon entries can be defined via formational features (e.g. a necessary condition to identify the gesture): handedness, hand shape, location, orientation, movement. Two gestures can be performed at the same time if the formational parameters are located on different dimensions.

The study of gesture requires relevant corpora and protocols. Kendon used video recordings collected in various cities during dinner parties, committee meetings, casual card games, interactions between customers and vendors at market stalls … (Kendon 2004). McNeill explored communicative gestures using narrative and descriptive protocols: retelling a story from a cartoon or comics (McNeill 1992), describing a house (McNeill, Quek et al. 2001), describing video vignettes showing small dolls interacting with simple objects (McNeill 2005). Butterworth and Beattie (Butterworth and Beattie 1978) examined films of tutorial sessions. Krauss asked subjects to describe pictures and asked actors to portray transcribed monologues (Krauss 1998).

Researchers in linguistics (McNeill 1992; Kita, van Gijn et al. 1998; Kipp 2004) established a set of criteria for temporally decoding movements. Some phases used for this temporal structure of gestures are (some of which can be optional):

- Preparation: the limb moves away from the rest position into the gesture space where it can begin the stroke
- Prestroke hold: a temporary cessation of movement before the stroke (possibly for repairing asynchrony of unfolding of manual and vocal movements)
- Stroke: phase with meaning and effort
- Stroke hold: stroke in the sense of meaning and effort but occurs with motionless hands (also called independent hold whereas prestroke and poststroke hold are called dependent holds)
- Poststroke hold: the hand freezes in the air before starting a retraction maintaining the stroke's final position and posture
- Retraction / partial retraction: the hands return to rest
- Beats: a repetitive phase (a number of repetitive movements where each movement would qualify as a stroke or a preparation)

[1] http://mcneilllab.uchicago.edu/pdfs/Coding_Manual.pdf

- Recoil: a small recoil movement that can happen after a forceful stroke where the hand lashes back from the stroke-end position

One of the key concepts that links gestures to the social interaction context is *space*. Space in interaction includes the areas defined by the orientation and the movements of the participants' bodies (Rodrigues, Kopp et al.). McNeill divided the gesture space into several areas (McNeill 1992). The *gesture space* is defined as a "shallow disk in front of the speaker" (McNeill 1992) where most gestures are performed. McNeill divides the space where gesture are produced into several areas (McNeill and Duncan 2000): the gesture space is divided in four regions (center-center, center, periphery and extreme periphery) and eleven coordinates (no coordinate, right, left, left-and-right (both hands), upper right, upper left, lower right, lower left, upper, lower, upper left-right, lower left-right).

Kipp et al. (Kipp, Neff et al. 2006) propose an annotation scheme for the spatial organization of hand/arm gestures. They focused on how to capture temporal structure and location information with relatively little annotation effort (hand height, hand distance, arm radial orientation and arm swivel). The coding scheme is illustrated using a virtual character in Figure 3

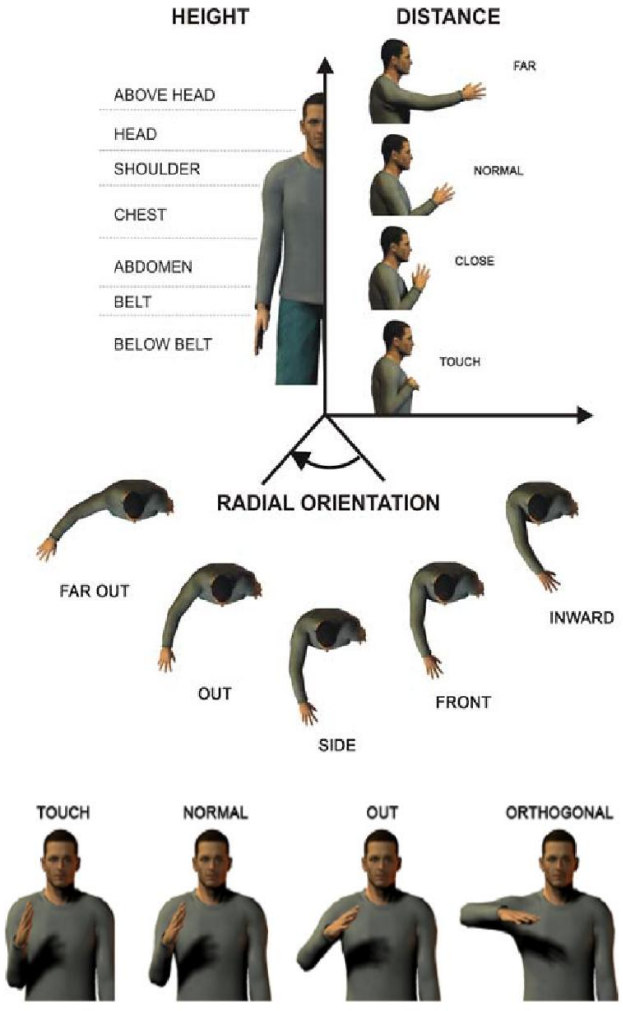


Figure 3: Annotation Scheme for Conversational Gesture (Kipp, Neff et al. 2006)

II.1.3 Symbolic descriptions of postures

Nonverbal studies have also considered the full body instead of focusing on the movements of a single body part such as the hand or the head.

Methods for describing full-body postures vary from grid-based observational studies to technical measures. Some researchers describe postures at a very global level. Argyle suggests three main posture categories (Argyle 1975): 1) standing, 2) sitting, squatting, and kneeling, and 3) lying. Harrigan et al. (Harrigan, Rosenthal et al. 2005) distinguish posture behaviors from action behaviors. Posture behaviors refers to overall postures (sitting, standing, lying), the frontal orientation of the trunk (facing, turned away), the trunk lean (forward, straight, backward, sideways), the arms and legs positions (folded arms, uncrossed legs) and the feet (flat on floor, under chair, on other knee).

In his symbolic Posture Scoring System, Bull defined a posture as a configuration of the body which is taken up and maintained for at least one second (Bull 1987). His coding system describes a posture in terms of a series of positions at the levels of the head, the arms, the trunk and the legs Figure 4. Head postures refer to an initial position in which the person is looking straight ahead in alignment with the direction of his chair, neither to right or left. Trunk postures refer to the upright position facing straight ahead in alignment with the direction of the chair, the trunk at 90 degrees to the chair seat. Arm postures refer to the position of the arm and the hand: whether it is touching the body, furniture or not touching anything. Leg postures include crossing the legs, moving the legs apart or together, drawing the legs back or stretching them out, and changing the orientation of the foot.

Bull also proposed a second system for coding the *dynamics* of the body. This second system describes both hand gestures and postures in terms of a series of movements rather than in terms of static positions.

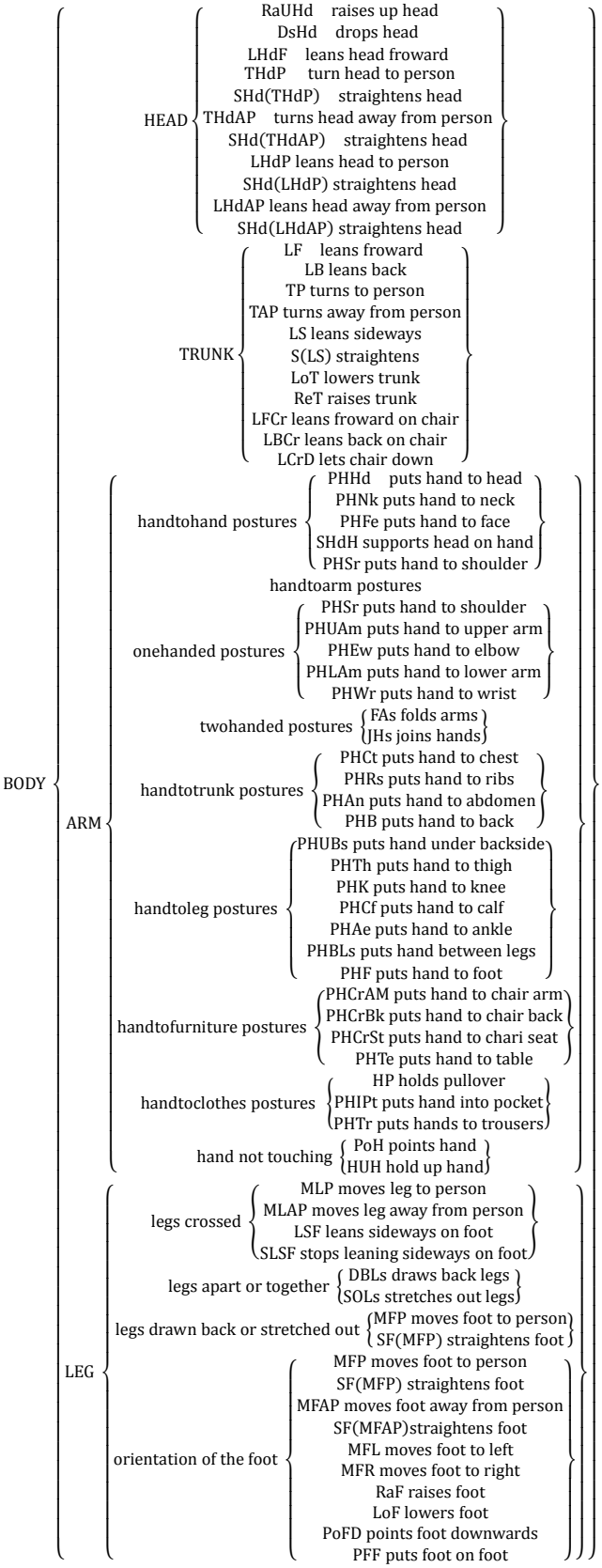


Figure 4: Posture Scoring System (Bull 1987)

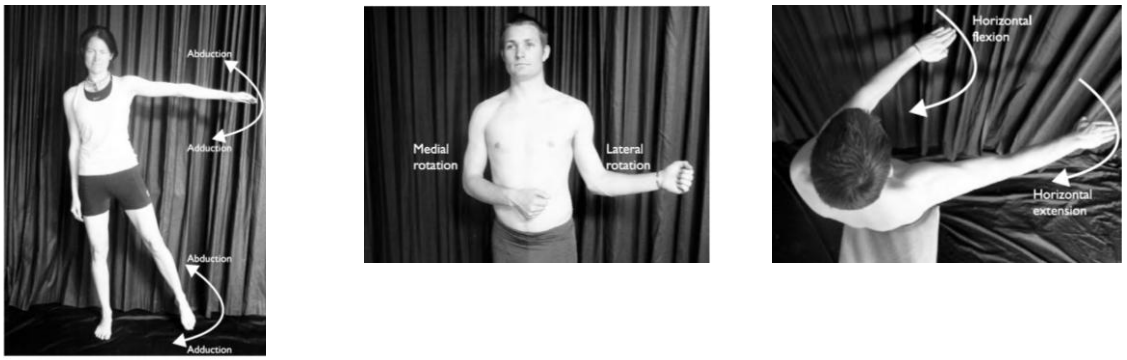
Coulson (Coulson 2004) considered the body as a system of interconnected rigid segments. Seven segments were defined for the upper body: head/neck; chest, abdomen, two shoulders/upper arms and two forearms). Six segments were defined for the lower body: two thighs, two shins and two feet. Fingers and toes were not considered. The joints and the rotations around the axes of the joints connecting the segments were used to describe body postures: head bend, chest bend, abdomen twist, shoulder adduct/abduct, shoulder swing and elbow bend, with seven degrees of freedom in total. The overall movement of the center of mass of the body was also included in terms of forwards, backwards and neutral.

Berthouze et al. (De Silva and Bianchi-Berthouze 2004) used a set of motion-capture-based spatial features for the descriptions of postures. 32 markers were used to provide motion information of the body. They constructed a corpus of 109 static expressive postures. The features involved the direction (19 features) and volume (5 features) of the body along the 3 orthogonal planes (horizontal, vertical and saggital). Each feature was normalized according to individual body characteristics (e.g. the distance between the right hand the left shoulder along the horizontal plane. They used Discriminant Analysis to select the discriminative features with respect to a model of emotional states.

Tan et al. (Tan, Slivovsky et al. 2001) used the Sensing Chair System (©Tekscan) to capture 50 postures (10 postures of 5 different types). Posture recognition was based on a Principal Component Analysis and Bayesian Classifier. Similarly, (D'Mello, Craig et al. 2008) used the Body Pressure Measurement System (©Tekscan) to construct a sitting posture corpus. The corpus contains two basic posture features: the back pressure, the seat pressure, the back change and the seat change.

Postures have also been studied and described in other disciplines. Anatomists (White 1991; Platzer and W. 2004) classify body motions as flexion vs. extension (the act of bending or straightening), abduction versus adduction (movement away from or toward the median plan), internal rotation versus external rotation (movement around an axis), and elevation versus depression (movement adjusting the height) (Figure 2). These classifications consider the act of performing body movements as well as the body position at the end of the movement.

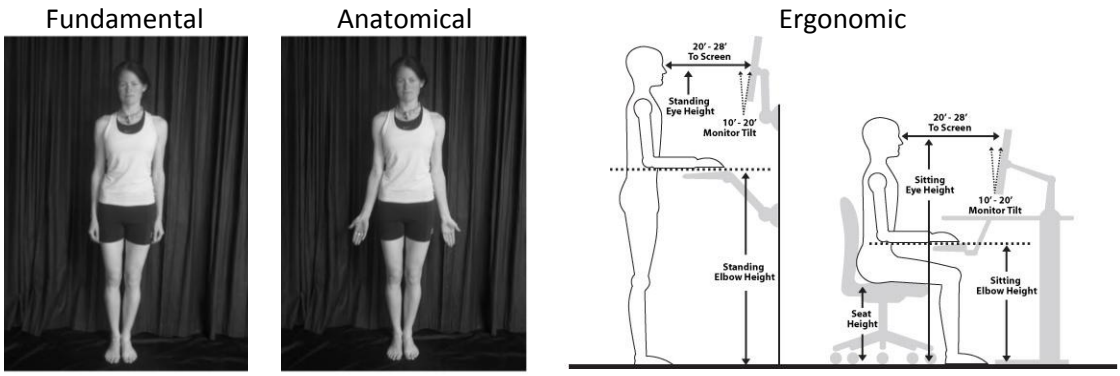
Figure 2: Body motion illustrations (Bartlett 1997)



Ergonomists (Corlett et al. 1986) define postures as conditions of the body. They use three methods to measure and describe postures using geometrical parameters: ordinal-scale-oriented methods, interval-scale-oriented methods and nominal-scale methods. Their aim is to find a trade-off between covering all possible rotational movements in any joint, and considering anatomic facts (e.g. a limited range of possible movements). For practical applications, the ordinal-scale-oriented methods are more commonly used than interval-scale-oriented methods. The nominal-scale methods place the description of postures according to posture typologies, which are generally profession-related.

Different disciplines consider the default posture in different ways. For example, the "fundamental" default posture is a 'stand to attention' position (Table 3 A), which is used in most of the disciplines such as psychology and animation design. The anatomists use the default posture in which the hand palms face forwards (Table 3 B). The ergonomists use a default posture (C) in which a person is either standing at a work station with her arms to be at a right angle when working, or sitting in front of the computer screen.

Table 3: Default postures in different disciplines: a) fundamental, b) anatomical, c) ergonomic



II.1.4 Convergence in dyadic conversation

Several experimental studies investigated bodily behaviors. In this section, we focus on one of the phenomenon that has been observed during conversations.

II.1.4.1 What is convergence?

(Condon and Osgton 1971) used film analysis to investigate movement coordination during interactions. They observed synchronous organizations of change between body motion and speech in both intra-individual and inter-individual behaviors: "The body of the speaker dances in time with his speech, and the body of the listener dances in rhythm with that of the speaker".

This phenomenon was considered at the beginning of the 80' as a simple effect of repeating the behavior of others, so called mimicry. In the 90', it was developed in the communication-accommodation theory (CAT) by Giles and Coupland (Giles, Coupland et al. 1991) where it goes beyond simple mimicry.

Convergence was defined as "a strategy whereby individuals adapt to each other's communicative behaviors in terms of a wide range of linguistic/prosodic/nonvocal features" (Giles, Coupland et al. 1991). The authors analyzed the ways through which individuals

converge or diverge their forms of speech styles in social interactions. Convergence is situated at the verbal level (mirroring vocabulary, accent, speech rate, grammar, voice etc) but also at the nonverbal level (matching other person's gestures, mannerisms, dress, hair, etc).

People may diverge their linguistic features but converge nonverbal behaviors, or vice-versa. (Giles and Johnson 1987) made the distinction between unimodal and multimodal convergent-divergent shifts, where the latter term implied shifting in several dimensions. It is likely that convergence of some features is matched by simultaneous divergence of others. For example, in a study of same and mixed-sex interactions, (Bilous and Krauss 1988) observed that females converged to males on some dimensions (total number of words uttered and interruptions) but diverged on others (laughter). Another study by (Scheflen 1964) observed that old friends or colleagues who have long-term ties converge their postures when they are temporarily arguing or taking opposing sides, as if to indicate the ultimate continuity of their relationship.

Evidence involving the *chameleon effect* extended the study of the convergence. It was found that people have a tendency to mimic other's posture, mannerisms and behaviors without awareness, so called the chameleon effect. The chameleon effect describe nonconscious mimicry of the postures, mannerisms, facial expressions and other behaviors of one's interaction partners, for example one's behavior passively and unintentionally changes to match that of others in one's current social environments (Lakin, Jefferis et al. 2003).

Postural convergence was observed in a study of (Charny 1966). The observations were based on a film of a psychotherapy session between a male therapist and a female patient. The annotations included converged postures and non-converged postures. The converged postures were either mirror-image converged postures (one person's left side is equivalent to the other person's right side) or identical converged postures (one person's left side is equivalent to the other person's left side). As shown in Figure 5, the pair of interlocutors in the foreground shows identical postures, while the pair of interlocutors in the background shows mirror-image postures (Bull 1987).

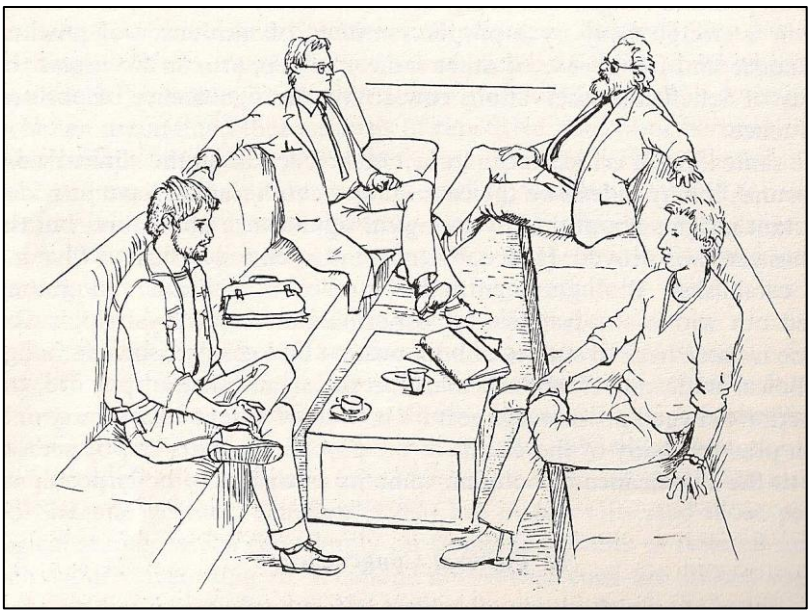


Figure 5: Identical postures (front) and mirror-image postures (back) (Bull 1987)

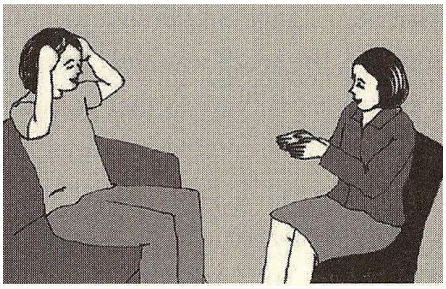
II.1.4.2 Why does convergence occur?

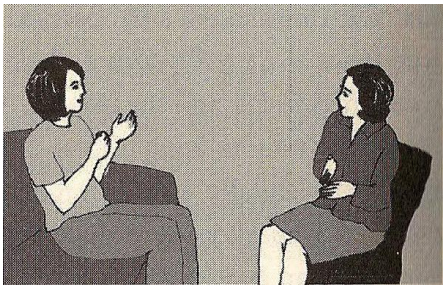
Converging the forms of nonverbal behaviors results from converging semantic representations. In the study of (Mol, Krahmer et al. 2011), it was found that perceiving vertical gestures of confederates increased viewers' production of vertical gestures (as well as in their use of one finger as an index). This means that viewers, when seeing vertical gestures would build the same semantic representation as in the gesture makers. The study suggested that the activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor (Garrod and Pickering 2004; Mol, Krahmer et al. 2011).

Converging gestures results from inhabiting some aspect of other's verbal thinking. This joint inhabitance can be observed in the form of gestural convergence. Kimbara (Kimbara 2006) studied gestural convergence based on a dyadic conversation between two friends. Her study suggested that gestural convergence was prominent when interlocutors are personally close. She described the phenomenon as a process of interpersonal synchrony. As shown in the two images on the top of Figure 6, the speaker on the right was expressing a gesture to describe the line of waiting passengers on Tokyo subway platforms during rush hour. The listener on the left is then preparing to perform the same gesture and move the hands to the gesture space. As shown in the two images on the bottom, the listener on the left is then mimicking the gesture.

II.1.4.3 Why do convergence and mimicry matter?

Research revealed that convergence and mimicry serve fundamental social functions. Mimickers can benefits from convergence. For examples, it was observed that when a waitress mimicked her customers, her tip amount significantly increased (van Baaren, Holland et al. 2003). Convergence also makes the interactor of the mimicker more confident and willing to respond in social interaction. Evidence from recent research in neurosciences showed that being imitated leads to positive feelings toward the imitator, as a specific neural area (called mOFC/vmPFC) is effectively connected to other areas during being imitated compared to not being imitated. The neuroscientists explained that being imitated leads to the activation of brain areas (mOFC/vmPFC) that are associated with emotion and reward processing (Kuhn, Muller et al. 2010).





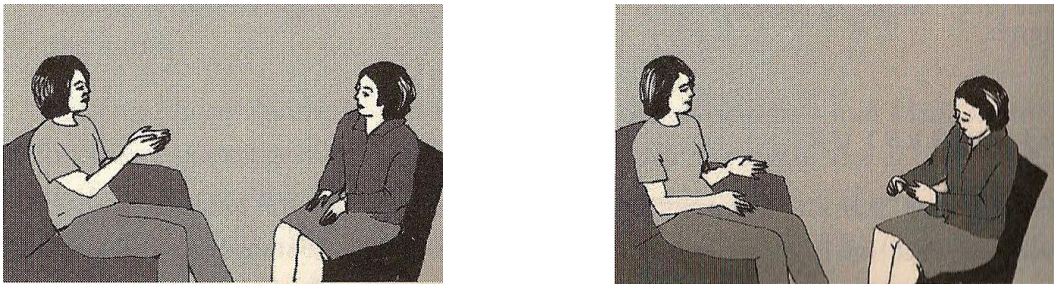


Figure 6: Gestural convergence observed in dyadic conversation (Kimbara 2002; McNeill 2005)

Convergence of facial expressions may also have a positive impact on the perception of interlocutors. By nine months, human infants begin to mimic emotional expressions, such as joy, sadness, and anger (Termine 1988). Mimicking facial expressions can result in actually adopting the motions and moods of others as well (Cacioppo, Petty et al. 1986). If we see or hear others laugh, we tend to laugh more ourselves (Young 1966).

Postural convergence is likely to increase rapport (LaFrance 1979; Lakin, Jefferis et al. 2003), liking (Lakin, Jefferis et al. 2003; Yabar, Johnston et al. 2006) and affiliation (Lakin, Jefferis et al. 2003) towards interlocutors. La France (LaFrance and Ickes 1981) found that students frequently displayed the same postural configuration as the posture displayed by the teacher. The extent of posture similarity was positively correlated with the student's ratings of: rapport, involvement, and togetherness. Scheflen (Scheflen 1964) posited that people in a group often mirror one another's posture and that this reflects a shared viewpoint. La France and Broadbent (LaFrance and Broadbent 1976) argued that posture mirroring is related to rapport. La France and Ickes (LaFrance and Ickes 1981) suggested that posture mirroring may not only reflect shared viewpoints and harmony but may actually be instrumental to achieving them. Several studies have found a relationship between behavior matching and self-reported rapport and involvement (Charny 1966; LaFrance 1979; Trout and Rosenfeld 1980). LaFrance (LaFrance 1979) suggested that the perception of rapport is drawn by the postural convergence that has causal priority over the perception of it. Chartran & Bargh (Chartrand 1999) found that matching behavior between interaction partners would increase liking and create a sense of smoother interactions.

Postural convergence also benefits from expressing empathy during social interaction. In an experimental study (van Baaren, Holland et al. 2003), an experimenter mimicked the posture of half of the participants, copying their body orientation, the position of their arms, and the position of their legs. The other half the participants was not imitated. After a couple of minutes, the experimenter pretended to drop his pen upon passing participants. The participants who had been imitated were significantly more likely to help the experimenter by picking up the pen than those who had not been imitated. Other studies suggest that postural convergence of nonverbal behaviors might increase pro-social behaviors and lead to other adaptive outcomes (efficiency, efficacy of communication) (Lakin, Jefferis et al. 2003; van Baaren, Holland et al. 2003).

Research on convergence revealed why this phenomenon occurs (resulting from the same representation), through which modalities it occurs (through verbal and nonverbal cues) and its impact on social interaction (promoting liking, rapport, presocial functions).

The current studies have two limits. Firstly, they use observation methods instead of statistical analyses. For example, we did not find any study in which convergence was formalised in such a way that it could be estimated using an algorithm. Existing studies code postures observed during convergence using categorical annotations (e.g. Non-congruent, identical or mirror-congruent (LaFrance and Broadbent 1976) and subjective judgments (van Baaren, Holland et al. 2003).

Secondly, these studies do not analyze the convergence of lower body postures (hips and lower limbs).

The notion of rapport has already been applied to virtual character, without a focus on the impact of full body postures (Gratch, Okhmatovskaia et al. 2006).

II.1.5 Corpora of postures during conversations: collection, annotation and analyses

Recent studies and projects using multimodal corpora address the question of multimodal annotation. Some corpora enable to study some features of postures during conversations.

Several corpora of dialogs and conversation exist. For example, the LUNA project (Raymond, Rodr ́guez et al. 2008) focused on spoken language understanding. The corpus was composed of human-machine and human-human dialogues. It proposed different levels of annotation, from morphosyntax to semantics and discourse analysis. SAMMIE (Kruijff-Korbayov, Kukina et al. 2008) is another project aiming at building multimodal resources in the context of human-machine interaction. Annotations were done using the (Carletta 2005). They mainly concern syntactic and discourse-level information. A comparable resource, also acquired following a Wizard-of-Oz technique, has been built by the DIME project (Pineda, Massa et al. 2002) for Spanish. These corpora of conversations do not feature postures.

Some parts of the body are sometimes annotated in corpora. For example, Daniel Loehr (Loehr 2004) annotated gestures and intonation in recordings of natural conversations. He did not observe that pitch and body parts (ex: legs) rise and fall together. He found that the apexes of gesture strokes and pitch accents aligned consistently. He found no significant correlation between gesture category and types (e.g. pitch accent, phrase tones). He found a rich rhythmic relationship in which hands, head and voice interplay and sometimes are synchronized and sometimes differ. After Loehr, Jannedy found that gestural phenomena are in robust co-occurrence with pitch accents in both laboratory and spontaneous speech (Jannedy and Mendoza-Denton 2005).

The Spontal project collected a Swedish corpus that contains 120 spontaneous dialogues in Swedish (at least 30 minutes each) (Edlund, Beskow et al. 2010). The two interlocutors were sitting face-to-face (captured in high-quality audio, high-resolution video and with a motion capture system). The Spontal corpus does not consider any higher-level manual annotation. The Spontal corpus contains motion capture data. Subjects are wearing Optitrack markers fixed to their torsos, arms and hands (hands, wrists, elbows, shoulders, sternum, and three

markers mounted on a head-worn tiara, as shown in Figure 7. However, the lower body (legs) was not tracked.



Figure 7: Subjects are wearing Optitrack markers fixed to torsos, arms and hands (hands, wrists, elbows, shoulders, sternum, and three markers mounted on a head-worn tiara (Edlund et al. 2010)

Few corpora and coding system enable to investigate the role of full body posture during spontaneous conversations. As we will see in the next chapter, several other corpora of bodily expressions have also been collected but for the specific study of emotion.

To sum up, existing nonverbal studies mainly considered the movements of a single body part such as the hand or the head. The static forms of the full-body received less attention. Although various methods for describing full-body postures already exist in Psychology, Ergonomics and Anatomy, the challenge remains with respect to find the tradeoff between the reliability of human observations and the accuracy of technical measures. Few corpora feature the annotation of lower body posture.

Furthermore postures, like any other nonverbal behavior cues, take place in a social context and are influenced by the norms which govern behavior both in the society at large and in an individual situation in particular. Hence, an understanding of the context should be the basis for understanding the meaning of body postures. However, few studies investigate how to code the forms of body postures under different contexts such as dyadic conversations and monologues.

II.2 Thesis contribution: the EXPO-sitting scheme for coding whole body static postures during seated dyadic conversations

II.2.1 Definition of a scheme for the manual annotation of whole-body postures

The previous annotation scheme for the CID corpus (I.4.1) only considered chest movements at the trunk level for coding bodily expressions (Bertrand et al., 2008). Few coding schemes of multimodal corpora consider the whole body posture. Particularly, lower body postures (e.g. legs) are seldom considered.

In order to extend the coding of whole body postures, we defined a new scheme for coding postures displayed in sitting situations met in the CID corpus. Our coding scheme is based on the Posture Scoring System (Bull, 1987) and the Annotation Scheme for Conversational Gestures (Kipp, Neff et al. 2006). Our coding scheme is called EXPO (ExPressive Postures).

We annotated a posture as "any configuration of bodily position which is taken up and maintained for at least one second" based on this definition of posture by Bull (Bull 1987) (Figure 8). Any movement that does not lead to a new posture should be excluded from the annotation of that posture. In the EXPO-sitting scheme, we annotate only the static part segment.

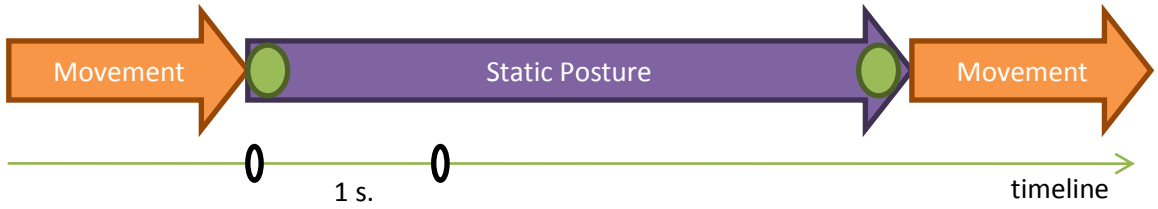


Figure 8: Segmentation of postures according to the Posture Scoring System (Bull 1987)

The EXPO-sitting scheme involves four body parts: arms, shoulders, trunk and legs. Since head movements have already been studies a lot (Ekman, Friesen et al. 2002; Quek, McNeill et al. 2002; Caridakis, Raouzaiou et al. 2006; Pianesi, Zancanaro et al. 2006), we decided to focus on the rest of the body. For example, compared to Bull, we added a specific attribute for the shoulders. With respect to the arm posture, Bull's system mainly distinguishes whether the hand touches or not another part of the body or an object. Kipp's scheme covers four additional spatial dimensions to code arm postures. We made a trade-off decision between the two systems: we kept the four dimensions from Kipp's coding scheme with respect to the height, distance, radial orientation and swivel degree of the arm, and created a new track describing the hand touching objects to consider suggestions from Bull's system. We also added two dimensions to describe respectively the arm posture in the sagittal plane and the palm orientation of the forearm and the hand. With respect to the leg posture, we added three new dimensions: the height of the feet, the orientation of the thigh and the way in which the legs are crossed (only for sitting positions). This scheme was implemented in the Anvil tool (Kipp 2001), the attributes and their possible values are shown in the following figure.

Figure 9. The EXPO-sitting posture coding scheme

$$\begin{bmatrix} \text{ARMS} & \begin{bmatrix} \text{ARMSHEIGHT} & \text{ArmHeight} \\ \text{ARMSDISTANCE} & \text{ArmDistance} \\ \text{ARMSRADIALORIENTATION} & \text{ArmRadialOrientation} \\ \text{ARMRADIALZ} & \text{ArmRadial Z} \\ \text{ARMSSWIVEL} & \text{ArmSwivel} \\ \text{FOREARMHANDORIENTATION} & \text{ForearmHandOrientation} \\ \text{ARMSTOUCH} & \text{ArmTouch} \end{bmatrix} \\ \text{SHOULDER} & [\text{SHOULDERTYPE} \quad \text{ShoulderType }] \\ \text{TRUNK} & [\text{TRUNKTYPE} \quad \text{TrunkType }] \\ \text{LEGS} & \begin{bmatrix} \text{LEGSHEIGHT} & \text{LegsHeight} \\ \text{LEGSDISTANCE} & \text{LegsDistance} \\ \text{LEGSSWIVEL} & \text{LegsSwivel} \\ \text{LEGSRADIALORIENTATION} & \text{LegsRadialOrientation} \\ \text{LEGTOTLEGDISTANCE} & \text{LegToLegDistance} \\ \text{CROSSEDLEGS} & \text{CrossedLegs} \end{bmatrix} \end{bmatrix}$$

$\text{ArmsHeight} = \{\text{above head, head, shoulder, chest, abdomen, waist, hip\&}buttock, thigh\}$

$\text{ArmsDistance} = \{\text{far, normal, close, touch}\}$

$\text{ArmRadialOrientation} = \{\text{behind, out, side, front, inward, inside}\}$

$\text{ArmRadial Z} = \{\text{forward, obverse, downward, reverse, backward, upward}\}$

$\text{ArmSwivel} = \{\text{touch, normal, out, orthoggonal, raised}\}$

$$\text{ForearmHandOrientation} = \left\{ \begin{matrix} \text{palm up,} \\ \text{palm down,} \\ \text{palm towards self,} \\ \text{palm away from self,} \\ \text{palm towards addressee,} \\ \text{palm away from addresse,} \\ \text{palm on side inwards,} \\ \text{palm on side outwards} \end{matrix} \right\}$$

$\text{ArmTouch} = \{\text{head, arm, trunk, leg, furniture, clothes, nottouching}\}$

$$\text{ShoulderType} = \left\{ \begin{matrix} \text{raise left shoulder,} \\ \text{raise right shoulder,} \\ \text{raise shoulders,} \\ \text{lower left shoulder,} \\ \text{lower right shoulder,} \\ \text{lower shoulders,} \\ \text{move shoulder forwars,} \\ \text{move shoulder back} \end{matrix} \right\}$$

$$\text{TrunkType} = \left\{ \begin{array}{c} \text{lean forward,} \\ \text{lean backward,} \\ \text{turn toward person,} \\ \text{turn away from person,} \\ \text{lean toward person,} \\ \text{lean away from person,} \\ \text{lower trunk,} \\ \text{raise trunk} \end{array} \right\}$$

$$LegsHeight = \{chest, abdomen, belt, buttock, thigh\}$$

$$LegsDistance = \{feet\ behind\ knee, feet\ in\ front\ of\ knee\}$$

$$LegsSwivel = \{feet\ outside\ knee, feet\ inside\ knee\}$$

$$LegsRadialOrientation = \{behind, out, side, front, inward\}$$

$$LegToLegDistance = \left\{ \begin{array}{c} knees\ apart\ ankles\ together, \\ knees\ together\ ankles\ apart, \\ knees\ together\ ankles\ together, \\ knees\ apart\ ankles\ apart \end{array} \right\}$$

$$CrossedLegs = \{ankle\ over\ thigh, at\ knees, at\ ankles, feet\ over\ feet, crossed\ legged\}$$

II.2.2 Posture annotations

Sixty minutes of the CID corpus involving a single pair of interlocutors (AB and CM) was annotated. The annotation yielded a total number of 760 postures (Table 20).

More arm postures were observed than leg postures and trunk postures as revealed by the number of annotated segments. The arm postures have an overall shorter duration than the leg postures. The subjects moved their arms more frequently than their legs. This difference is probably due to the conversational context: the two subjects are both sitting on a chair with a back. This might explain why they move their arms more often than their trunk or their legs.

The number of segments in the different body parts and their durations are similar for the two subjects AB and CM.

Table 4: Postures annotated for different body parts in the CID corpus (video AB_CM): number of annotated segments and their total duration in seconds

	Subject AB		Subject CM	
	duration (s)	Nb. of annotated segments	duration (s)	Nb. of annotated segments
Left Arm	2840	114	2700	151
Right Arm	2723	129	2764	120
Shoulder	2450	37	2579	29
Trunk	2563	59	3372	38
Left Leg	3387	56	3478	26
Right Leg	3401	20	3487	15
TOTAL	17364	415	18379	379

II.2.3 Clustering postures

II.2.3.1 Research goal

The EXPO-sitting coding scheme leads to annotating separately the positions of the different body parts. Thus each posture annotations contain descriptions concerning only one aspect of the body posture. For instance, the trunk is leaning forward, or the left arm is rising at the level of shoulders. The posture observation is thus represented as a vector of monotonic posture annotations performed by different parts of the body.

We were willing to investigate the most common whole-body postures that the subjects take up, so called posture type. The posture type would refer to a typical posture generalized based on the posture observations.

II.2.3.2 Data extraction

We exported 19183 frames from Anvil for the posture annotations of 15 minutes of only one subject (CM). Then, we converted these 19183 instances into a set of significant postures according to the two criteria below:

- Exclude the frames in which there is no annotation in any track
- Exclude the successive frames that contain exactly the same annotations

These criteria are inspired by those of the Posture Scoring System (Bull 1987). The author emphasized that if the speaker moves from one posture without establishing a different posture and then returns to the original posture, the time spent moving should be excluded from the total time length of that posture. For that reason, we excluded the frame in which there was any movement in any body parts.

The extracted data contained a matrix of 89 frames by 25 features. The 25 features were composed of 24 attributes from the EXPO coding scheme and one feature of posture type that describe the recurrent posture to be found.

II.2.3.3 Data analyses

We applied a Hierarchic Clustering on the extracted data. This method enables to make a global-view analysis on the posture annotations and cluster different posture combinations into similar form categories. We created a Hierarchical cluster tree using the Euclidean distance metric and Ward's linkage algorithm.

According to the Euclidean distance, if p = (p_1, p_2) and q = (q_1, q_2), then the distance between p and q is calculated as:

$$d(p,q) = \sqrt{\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}}$$

We used Ward's linkage algorithm (minimum variance), because it is the only appropriate linkage algorithm for Euclidean distances.

The formula linkage uses the following variables:

1) cluster r (or s) is formed from clusters p and q;

2) n_r (or n_s) is the number of elements in cluster r (or s);

3) x_{ri} is the i^{th} object in cluster r, $\overline{x_r}$ (or $\overline{x_s}$) is the centroid of cluster r (or s).

The distance between the clusters r and s is described as follows:

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r + n_s)}} \|\overline{x_r} - \overline{x_s}\|_2$$

Three gross body posture types were found for one of the speakers (CM) (during the first 15 minutes of the conversation) (Figure 10). Each U-shape line represents the distance between the two posture combinations. The branches are grouped according to distance. Each point on the X axis represents one instance (out of the 97 postures).

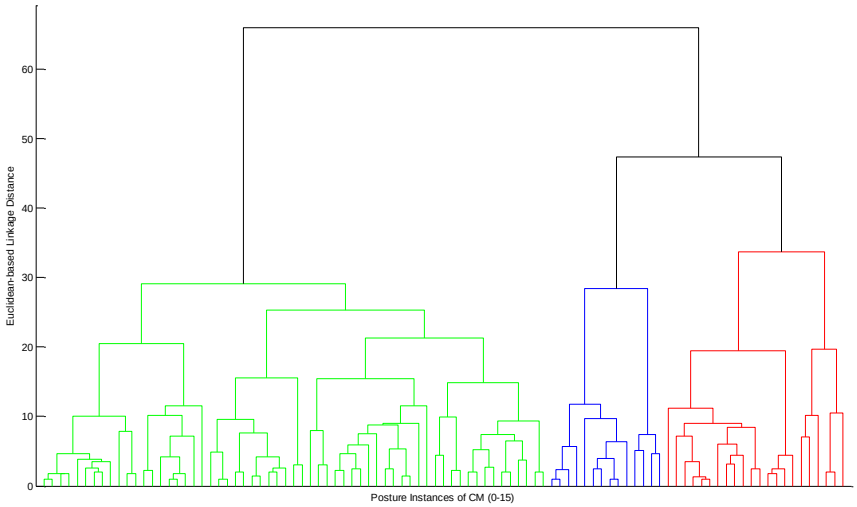


Figure 10: Dentogram of the clusters based on the 89 extracted posture instances (each color represent a different posture type)

The 97 posture instances were then plotted in a 2-dimensional space using a Principal Component Analysis (PCA). PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The PCA generates a new set of features (called principal components). Each principal component is a linear combination of the original features. We selected a PCA since it might enable us to discover the postures that are the most frequently taken up in the corpus. As shown in the figure below, the red points refer to posture type #1, which represents 62.89% of the 97 posture instances. The green points refer to posture type #2, which represents 14.43% of the 97 posture instances. The blue points refer to posture type #3, which represents 22.68% of the 97 postural instances. The X and Y axes represent the first two principal components that explained roughly 51.02% of the total variance.

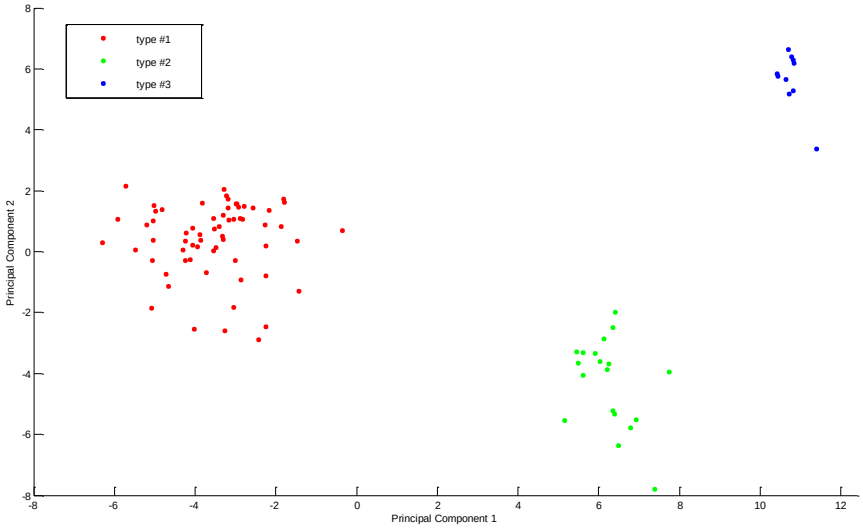
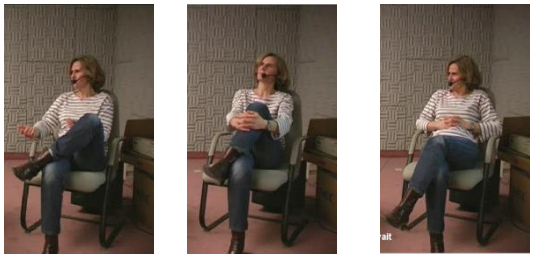


Figure 11: Clustering of posture types

As seen in Figure 12, we selected three representative frames to illustrate the most common postures.



posture type #1 | posture type #2 | posture type #3

Figure 12: Illustrations of three typical leg postures occurring during the first 15 minutes of conversation for subject CM.

Figure 13 shows that the first six principal components explain 80.93% of the total variance. This method provides a reasonable way to reduce the dimensions in order to visualize the

clustering result. The PCA also computed scores for each of the 28 attributes that map into the space of the 6 principal components, as shown in Table 5. The attributes that received the highest absolute values by principal components include (in bold in Table 5):

- Arm height (for both arms),
- Arm distance (for both arms),
- Arm touch (for both arms),
- Left leg height (for both legs),
- Leg radial orientation (for both legs),
- Leg-to-leg distance (for both legs),
- Right leg crossed legs (the way in which the right leg is crossed with the other leg).

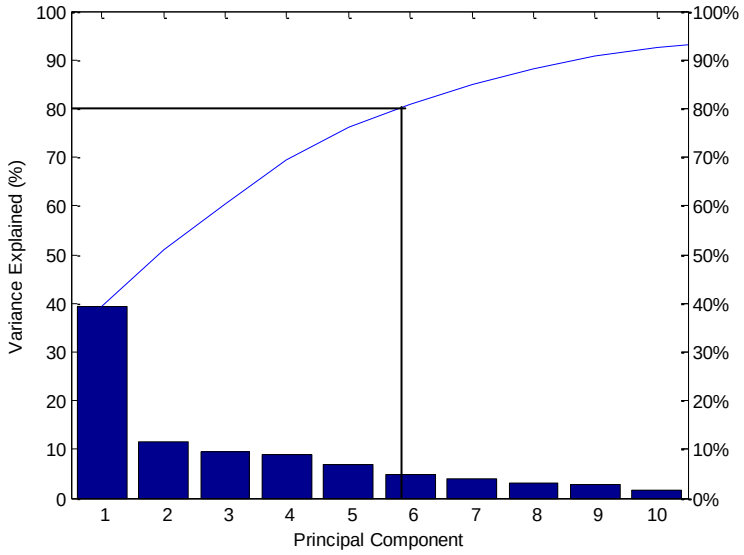


Figure 13: Percentage of explanations by the principal components

Table 5: PCA scores: transforming the original data of leg posture instances into the space of the first 6 principal components

28 Attributes	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
LeftArm:ArmsHeight	**-0.4675**	0.1645	**-0.2022**	-0.0213	0.1055	0.0682
LeftArm:ArmsDistance	-0.2546	0.0858	-0.0782	**-0.3245**	0.1554	-0.1364
LeftArm:ArmsRadialOrientation	-0.389	0.0963	-0.1397	0.0664	0.1235	-0.1391
LeftArm:ArmsRadialZ	-0.1667	0.0591	0.0238	0.0424	0.0517	-0.0089
LeftArm:ArmsSwivel	-0.2451	0.1156	0.0534	-0.2367	0.0901	-0.0535
LeftArm:ForearmHandOrientation	-0.2965	0.1154	-0.1264	-0.0969	-0.0321	-0.259
LeftArm:ArmsTouch	**-0.4458**	0.2074	0.019	0.2502	-0.0892	**0.4932**
RightArm:ArmsHeight	-0.2188	**-0.4782**	0.0104	-0.0354	0.0486	-0.2549
RightArm:ArmsDistance	-0.1092	-0.2361	-0.0003	**-0.3594**	0.0162	-0.0642
RightArm:ArmsRadialOrientation	-0.1499	-0.3608	0.1877	0.1581	-0.1567	-0.2104
RightArm:ArmsRadialZ	-0.0679	-0.1224	-0.0098	0.1442	-0.0403	-0.0062
RightArm:ArmsSwivel	-0.1116	-0.219	0.0354	-0.1829	-0.0182	0.036
RightArm:ForearmHandOrientation	-0.1489	-0.245	0.0679	-0.0116	-0.1992	-0.2675
RightArm:ArmsTouch	-0.2325	**-0.3748**	0.1415	**0.3591**	-0.2236	**0.3301**
Shoulder	-0.0137	0.0058	-0.0074	0.0276	-0.0038	-0.0274
Trunk	-0.0184	0.0554	-0.0178	0.1266	-0.2054	-0.152
LeftLeg:LegsHeight	-0.0687	0.0712	**0.527**	-0.2856	-0.027	0.1202
LeftLeg:LegsDistance	-0.0172	0.1137	0.0081	-0.0178	-0.1886	-0.0956
LeftLeg:LegsSwivel	-0.0483	0.1261	0.2089	-0.0769	-0.1922	-0.0042
LeftLeg:LegsRadialOrientation	-0.0664	**0.2159**	**0.4919**	-0.1644	**-0.3308**	-0.0346
LeftLeg:LegToLegDistance	**-0.0053**	0.0729	0.0497	0.0497	-0.1416	-0.089
LeftLeg:CrossedLegs	-0.0264	0.0821	0.1531	-0.1186	-0.1311	0.0101
RightLeg:LegsHeight	-0.0285	0.1786	0.2197	**0.3401**	0.179	**-0.3267**
RightLeg:LegsDistance	-0.0073	0.0364	0.0452	0.0202	0.0476	-0.0693
RightLeg:LegsSwivel	-0.0098	0.043	0.0538	-0.0163	0.0655	-0.0842
RightLeg:LegsRadialOrientation	-0.0294	0.1266	0.2769	0.3041	**0.2997**	-0.2586
RightLeg:LegToLegDistance	**-0.0008**	**0.2376**	**-0.1923**	0.2006	**-0.4531**	**-0.3267**
RightLeg:CrossedLegs	-0.0194	-0.0434	0.3096	0.1552	**0.4637**	0.0013

II.2.3.4 Discussion

Our results are in line with a previous study (Bressem 2007), which supports that during dyadic conversations speakers actually use only a small set of shapes and orientations of gestures over a large variety of forms that they could use. For example, the study by Bressem identified 6 recurrent hand forms over 32 possible hand forms. It also identified 5 recurrent hand orientations out of 13 possible orientations of the hand. Thus, this study observed that German speakers use specific hand shapes, orientations of the hand, movement patterns and positions in space recurrently during dyadic conversation. The results of our clustering process suggest similarly that subject CM uses mostly 3 types of postures, possibly the ones which are the most appropriate in the physical and social context.

II.2.4 A scheme for coding gesture space[2]

II.2.4.1 Annotations

The focus of this thesis is not only on posture but also on how the space is used during bodily interaction. We did not find any multimodal corpus approach defining and applying a scheme for manually annotating the gesture space during conversations. We thus defined our own scheme for coding the gesture space based on the textual and graphical descriptions of the gesture space proposed by (McNeill and Duncan 2000). The gesture space is divided in four regions (center-center, center, periphery and extreme periphery) and eleven coordinates (no coordinate, right, left, left-and-right (both hands), upper right, upper left, lower right, lower left, upper, lower, upper left-right, lower left-right). We used McNeill's diagram for defining our coding scheme (Figure 14). These attributes were annotated independently for the two hands. The left-and-right coordinate is used to code a gesture produced with both hands.

[2] The study described in this section was conducted in collaboration with several partners from the OTIM project including Gaëlle Ferré and Marion Tellier

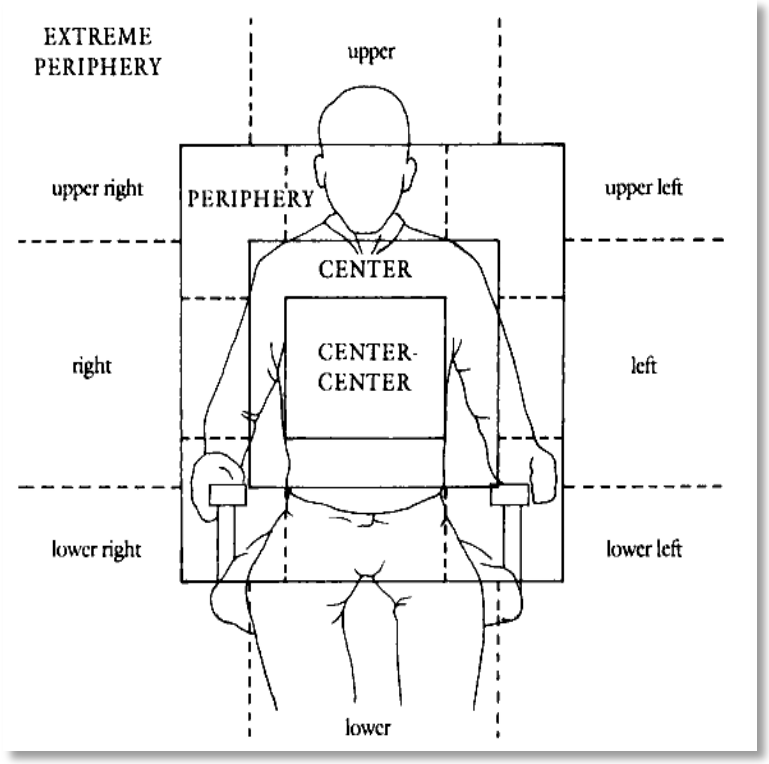


Figure 14: Gesture space (McNeill 1992)

Annotations of region and coordinates are based on the point of maximum extension of the gesture. When annotating the GestureRegion, the gesture that occurred in the axis of the armrests of the chair is judged as being in the periphery of the body. When the gesture is produced outside of this axis, we annotate the spatial area as being the extreme periphery. The landmark is thus the position of the wrist.

II.2.4.2 Methods

In order to validate the coding scheme, we recruited three coders to annotate the GestureSpace including GestureRegion and GestureCoordinates on three minutes of one video of the CID corpus (a segment of the video AB-CM: 0-15).

The intercoder agreement is the degree of agreement among coders. It gives a score of how much homogeneity, or consensus, there is in the ratings provided by coders. If various coders do not agree, either the scale is defective or the coders need to be re-trained.

Measures of intercoder agreement need to be insensitive to coders' bias and differences in the presented items of different types. They need to enable separate analysis of accuracy for each item type and a comparison of performance between studies even with different numbers of categories (Cavicchio and Poesio 2009). There are a number of statistics that meet these requirements, such as, joint-probability of agreement, Cohen's kappa and the related Fleiss' kappa, inter-rater correlation, concordance correlation coefficient and intra-class correlation. Different statistics are appropriate for different types of measurement.

Kappa is widely used to measure the degree of agreement between two independent judges. Extension of kappa to three or more judges has traditionally involved measuring pairwise agreement among all possible pairs of judges.

Kappa is commonly used to assess the intercoder agreement in multimodal corpora. The kappa coefficient k measures pairwise agreement among two coders making category judgments. It is generally thought to be a more robust measure than simple percent agreement calculation since it takes the observed categories' frequencies as given, and corrects for chance.

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Figure 15: Cohen's kappa

II.2.4.3 Results

A strong agreement was observed for Gesture Region: 0.65 for the right hand, and 0.67 for the left hand.

However, a low agreement was observed for the coding of the Gesture Coordinates: 0.26 for the right hand, and 0.59 for the left hand.

Table 6: Average kappa values for each categorical item of GestureSpace for three coders

Kappa (k)	Gesture Region	Gesture Coordinates
Right Hand	0,65	0,26
Left Hand	0,67	0,59

II.2.4.4 Discussion

This low agreement on how to code Gesture Coordinates might be due to several factors. Firstly, the number of categories is high (12 categories). Secondly, there is some interaction between Gesture Region and Gesture Coordinates. If the coders disagree about how they code the Gesture Region, they are also likely to annotate Gesture Coordinates in a different way. For instance, according to our scheme (and to McNeill schema) a "no coordinate" should be given for a gesture produced in the center-center region. All other gesture regions subsume some coordinate values. This means that whenever coders disagree between the center-center or center region, the annotation of the coordinates cannot be congruent either.

Several solutions might be investigated in the future to cope with this low agreement. The number of possible values for the Gesture Coordinates attributes could be reduced. Motion capture might be used to complement or validate the manual annotation of gesture space.

II.2.5 Conclusion

In this section II.2 we achieved the following goals: first, we proposed a scheme (called EXPO-sitting scheme) for coding conversational expressive postures. This coding scheme integrates features of the Posture Scoring System by Bull (Bull 1987) and the Annotation

Scheme for Conversational Gestures by Kipp (Kipp, Neff et al. 2006). It enables to code the positions of different body parts including arms, trunk, shoulders and legs. 26 features are associated to the EXPO-sitting scheme.

Second, we used a Principal Component Analysis to reduce the dimensions of the coding scheme, so that a trade-off was found with respect to the integrality of the whole-body descriptive features and the relevance to the corpus being investigated. Over the 26 features described in the EXPO-sitting scheme, seven features were found to be relevant to describe the sitting postures that occur during a one-hour dyadic conversation by two French speakers. These seven features include arm height, arm distance, arm touch, left leg height, leg radial orientation, leg-to-leg distance and the ways in which the right leg crossed with the left leg.

Third, we applied a clustering to the 97 postures displayed by one subject that were extracted from the annotations of a 15mn video. We aimed to reveal the postures that were taken up the most frequently. Our assumption was that only a small set of postures would be used under specific context such as a dyadic conversation. We found indeed that three postures were mainly observed.

II.3 **Thesis contribution: Estimating postural convergence during dyadic conversations**

The section on related work explained why the phenomenon of convergence might be of interest for the study of postures and space. Yet, we did not find any adequate corpus and coding schemes to enable such studies. This section explains how we exploited the CID video corpus and how we defined coding schemes that we applied for studying postural convergence.

II.3.1 A formal definition of postural convergence

As explained previously, the number of postural segments and their durations are similar for the two subjects AB and CM. We were willing to study whether this might be explained by a phenomenon of postural convergence between the two subjects.

During a preliminary observation of the video, we found several cases during which the two subjects display similar postures, as illustrated in Figure 12. In the first image (top left), the speaker raises her left arm and put it on the top of her head. Half a second later, as shown in the second image (top right), the listener is preparing to put her right arm on her head, and this position is maintained for more than one second. This is an illustration of what we call a postural convergence.



Figure 16: An example of postural convergence for a pair of interlocutors in CID (See text for explanations)

Our formal definition of postural convergence is inspired by the cross-lag panel technique (Kenny 1975):

> There is **postural convergence** when the posture of the listener at time t adapts to the posture of the speaker at time t-1.

We are thus willing to investigate whether the posture displayed by the speaker at time t would be influential in establishing the posture of the listener at time t+1. We study this phenomenon that is likely to occur in a *shared space* between two subjects (Figure 17) (McNeill 2005).

This notion of *shared space* is different from the notion of *gesture space* that we studied in the previous section.

The gesture space refers to the individual space, which is a body centered area that falls in front of the body. McNeill related the different gestures to different regions of the gesture space. For example, deicitic gestures are related to the periphery, iconic gestures are related to the center-center space and the metaphoric gestures are related to the lower center. However, the gesture space does not concern the lower body and it focuses on the gestures that an individual can produce, even when being alone (Battersby 2011).

In this section, we are interested in the union of the individual spaces of the two subjects where the phenomenon of postural convergence might occur (Özyürek 2002). We used the term of convergence rather than mimicry, as convergence not only explains the repetitions of the forms but also the activations of the same semantic representations.



Figure 17: Spatial arrangement of subject AB and subject CM and the space they share during a conversation

II.3.2 Relations between the speaker / listener role and the postures

II.3.2.1 Data collection

Our definition of postural converge requires to have some information about the speaking role of each subject (e.g. speaker or listener). Such annotations will enable us to investigate whether convergence occurs in the postural expressions taken up by two subjects during a conversation.

Thus, we annotated one hour of the CID corpus in terms of speaker / listener roles to investigate the convergence between both. The speaker / listener role is based on the concept of floor, which refers to the right of one member to speak in preference to other members (Bavelas, Coates et al. 2000). We focus only on this notion of speaking role and do not consider more sophisticated functions such as backchannels (the verbal and nonverbal listener responses) or turn-taking (the turn construction and the turn allocation) (Sacks, Schegloff et al. 1974; Allwood 1999). Our hypothesis here is that the speaker /listener role are already enough to investigate postural convergence. Thus, the backchannel behaviors occurring during a listening turn were not coded in the annotations that we describe below (i.e. they were included in a segment annotated as listener).

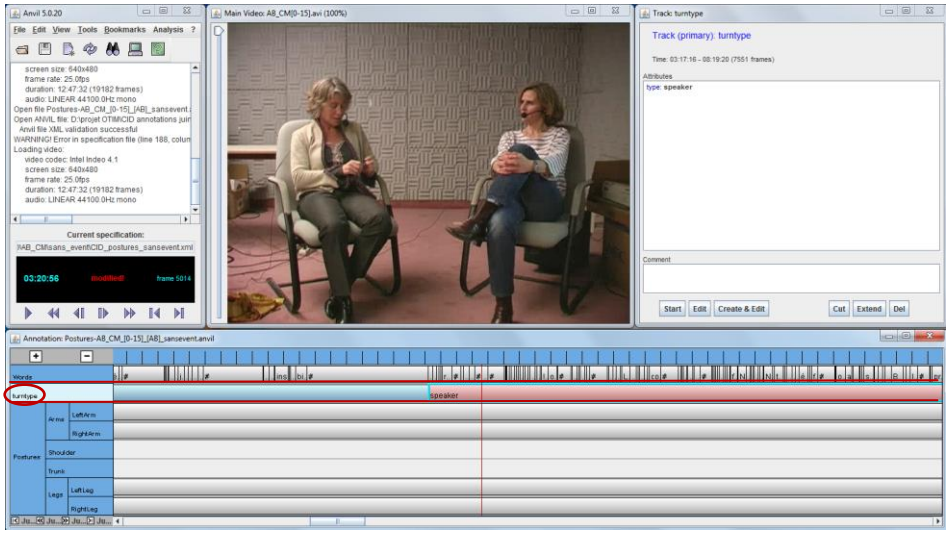


Figure 18: Annotations of the speaker / listener role using Anvil (Kipp 03)

During this one-hour conversation, a total of 33 segments was annotated with respect to the speaker / listener role (16 segments for subject AB and 17 segments for subject CM). The segments of both subjects were also almost in balanced in terms of duration (Table 7).

Table 7: Duration of annotated postures according to the conversation role: in terms of number of frames and seconds (25 frames per second)

	AB		CM		TOTAL	
	Nb. frames	Duration (sec.)	Nb. frames	Duration (sec.)	Nb. frames	Duration (sec.)
Speaker	32018	1281	27590	1104	59608	2384
Listener	25293	1012	33391	1336	58684	2347

We then analyzed separately the different body parts:

1) the relations between the **upper** body posture and the speaker / listener role, and

2) the relations between the **lower** body posture and the speaker / listener role.

II.3.2.2 Association analysis with respect to the upper body

We observed several recurrent patterns that distinguish the speaker and the listener roles in terms of body postures. We performed an association analysis to determine if there are nonrandom associations between these two categorical variables, i.e. the speaker / listener role and each of the posture features.

Instead of a chi-square test, we chose the Fisher's Exact Test for the association analysis on our data, because 1) it is relevant when the data set has a small sample size without being affected by a lack of continuity; 2) this test takes into account a n*m contingency table, instead of 2*2 or 3*3.

We exported from Anvil a set of 23 contingency tables as a function of different combinations between the speaker / listener role and the 28 posture features (5 features did not lead to compute contingency tables in relation to the feature of the speaker / listener role).

In Table 8, the p-values are given for each of the combinations involving the features of the upper body. If Sig=1 ($p<.05$), then the measured variables are significantly associated. If Sig=0 ($p>.05$), then the measured variables are not associated.

The results confirmed that there was a statistically significant association between the two measured variables. The feature of the speaker / listener role is strongly associated to each of the listed attributes of the arm postures, i.e. arm distance, arm height, arm radial orientation, arm swivel, arm touching and forearm orientation.

Table 8: Measures of association between the speaker / listener role and the attributes of the upper body

Attributes	Sig.	P-Value
left_arm_distance	1	.0001
left_arm_height	1	0
left_arm_radial_orientation	1	0
left_arm_swivel	1	.0001
left_arm_touching	1	0
left_foream orientation	1	0
right_arm_distance	1	0
right_arm_height	1	0
right_arm_radial_orientation	1	0
right_arm_swivel	1	.0002
right_arm_touching	1	0
right_foream orientation	1	.0001

We then performed a qualitative analysis based on the features that have association with the feature of the speaker / listener role. We discuss below these features according to which strong differences between speakers and listeners (displayed in bold in the tables below) were observed in terms of percentages.

In terms of arm radial orientation, speakers are more likely to direct the arms forward (61.69%) compared to listeners (57.05%), and direct to the side (20.97%) compared to listeners (11.41%). In contrast, listeners tend to direct the arms inward (11.41%) more frequently compared to speakers (8.87%) and inside (17.45%) compared to speakers (4.44%).

Table 9: Arm radial orientation

	Front	Side	Inward	Inside	Behind	Far/Out of
speaker	**61,69%**	**20,97%**	8,87%	4,44%	2,42%	1,61%
listener	57,05%	11,41%	**11,41%**	**17,45%**	2,01%	0,67%

In terms of arm swivel, speakers are more likely to keep the arms far from the body when elbows are brought in tight to the body (12%) compared to listeners (5.84%). Listeners are more likely to maintain arms close or touched to the body (47.40% and 43.51% respectively) compared to speakers (42% and 39.6% respectively). This finding suggests that speakers tend to display large postures when having the floor.

Table 10: Arm swivel

	Out	Touch	Normal (close)	Orthogonal	Raised
speaker	**12,00%**	42,00%	39,60%	0,40%	6,00%
listener	5,84%	**47,40%**	**43,51%**	1,30%	1,95%

In terms of arm touching, speakers are more likely to not touch their body (22.44%) compared to listeners (4.55%). In contrast, listeners are more likely to touch their trunk (15.58%) and clothes (34.42%), compared to speakers (6.3% and 27.17% respectively).

Table 11: Arm touching

	Not touching	Clothes	Trunk	Head	Arm	Leg
speaker	22.44%	27.17%	6.3%	9.84%	21.65%	12.60%
listener	**4.55%**	**34.42%**	**15.58%**	5.19%	24.68%	15.58%

In terms of arm height, the results showed that speakers tend to put their hands at the level of their head (11.07%) and shoulders (5.93%) compared to listeners (4.55% and 0% respectively). In contrast, listeners tend to put hands at the height of chest (17.53%) and abdomen (50%) compared to speakers (12.25% and 36.76% respectively).

Table 12: Arm height

	Above head	Head	Shoulder	Chest	Abdomen	Waist	Hip/buttock
speaker	1,58 %	**11,07%**	**5,93 %**	12,25 %	36,76 %	26,48 %	5,93 %
listener	0,65 %	4,55 %	0,00 %	**17,53 %**	**50,00 %**	25,97 %	1,30 %

In terms of arm distance, speakers tend to keep a short distance between the wrist and the median plan (26.59%) compared to listeners (9.8%). Listeners tend to keep extreme distance positions by either having the elbows touch the body (33.33%) or put the arms far straight (15.03%), compared to speakers (20.24% and 9.13% respectively).

Table 13: Arm distance

	Far	Short/Normal	Close	Touch	Backup
speaker	9,13%	**26,59%**	40,87%	20,24%	3,17%
listener	**15,03%**	9,80%	41,18%	**33,33%**	0,65%

Finally, in terms of the orientation of forearms, speakers tend to put the palm of forearm up (7.66%) compared to listeners (0.67%). In contrast, listeners tend to put it towards self (66.67%) compared to speakers (59.68%).

	Palm up	Palm down	Palm towards self	Palm away from self	Palm towards addressee	Palm on side inwards
speaker	**7,66%**	19,76%	59,68%	3,23%	0,40%	9,27%
listener	0,67%	16,00%	**66,67%**	4,00%	0,00%	12,67%

The main differences between speakers and listeners in terms of postures at the level of the upper body are illustrated in the figure below.

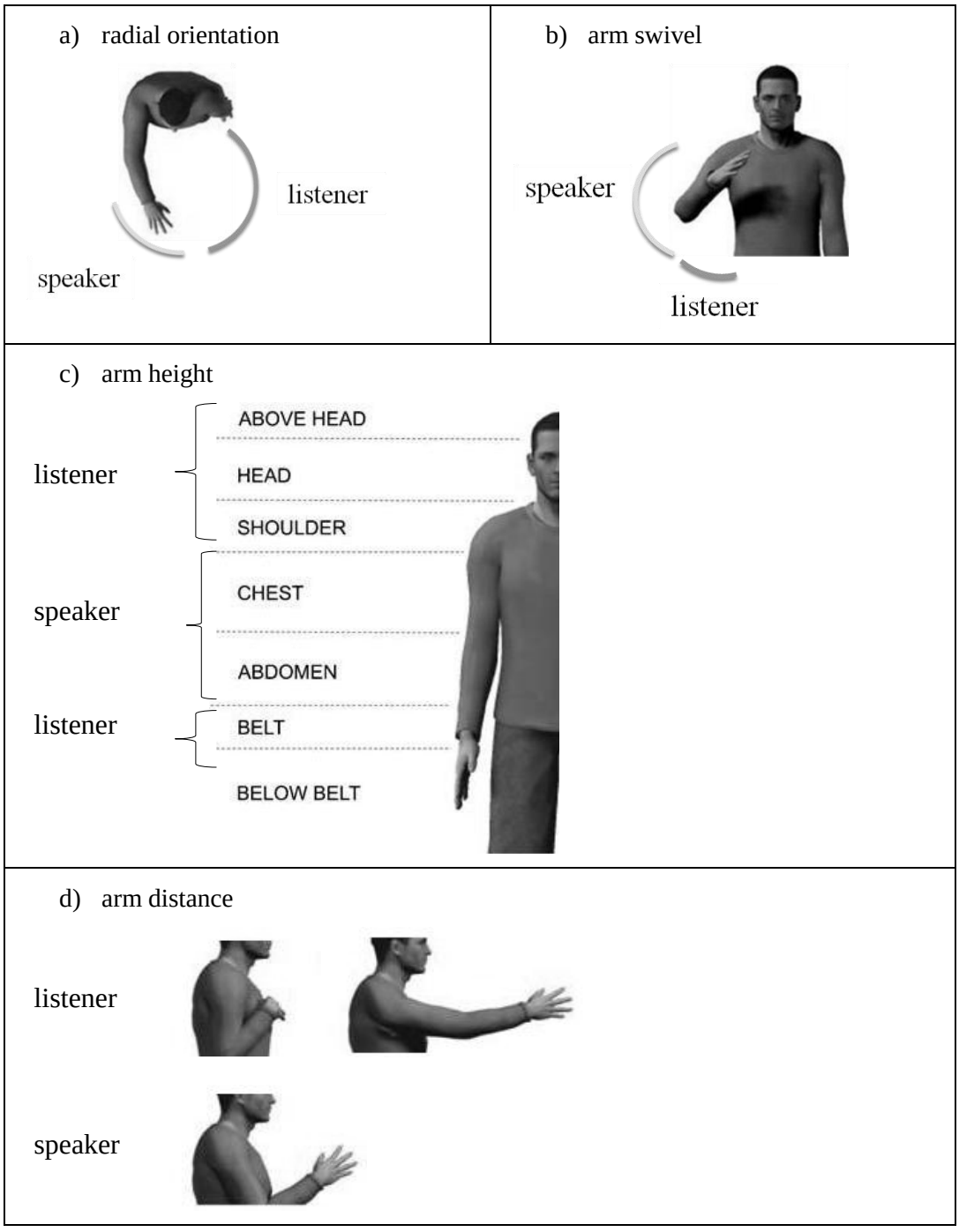


Figure 19 : Principal arm configurations according to speaker / listener roles

II.3.2.3 Association analysis with respect to the lower body

We also measured the association between the speaking role and the lower body postures.

The p-values for each of the combinations involving the features of the lower body are listed in Table 13. The results confirm that there is a statistically significant association between the

two variables. The speaker / listener role are strongly associated to each of the listed attributes of the leg postures, crossed legs, leg distance, leg-to-leg distance, leg radial orientation, leg swivel (only for the right leg).

Table 14: Association between the speaker / listener role and the attributes of the lower body (significative results in bold)

attributes	Sig.	P-Value
left_leg_crossedlegs	**1**	**.0001**
left_leg_distance	**1**	**.0281**
left_leg_height	0	.0649
left_leg_legtolegdistance	**1**	**.0017**
left_leg_radialorientation	**1**	**.0033**
left_leg_swivel	0	.0706
right_leg_crossedlegs	**1**	**.0034**
right_leg_distance	**1**	**.0493**
right_leg_legtolegdistance	**1**	**.0092**
right_leg_radialorientation	**1**	**.0233**
right_leg_swivel	**1**	**.0078**

Similarly to the analyses described earlier for the upper body, we then performed a qualitative analysis based on the features of the lower body that have an association with the feature of the speaker / listener role. We discuss below these features according to which there are the strongest differences between speakers and listeners (displayed in bold in the tables below).

In terms of leg-to-leg distance, speakers are more likely to display postures in which the ankles are kept together whenever knees are kept together (6.12%) or not (38.78%), compared respectively to listeners (2.13% and 14.89%). In contrast, listeners are more likely to display postures in which they keep the ankles apart. This is true whatever the configuration of the knees: knees apart (38.30%) or knees together (44.68%).

Table 15: Leg-to-leg distance

	Knees apart + Ankles together	Knees together + Ankles apart	Knees together + Ankles together	Knees apart + Ankles apart
speaker	**38,78%**	34,69%	**6,12%**	20,41%
listener	14,89%	**44,68%**	2,13%	**38,30%**

These results concerning leg-to-leg distance are in line with those concerning the leg-crossing forms (Figure 20 **a**). Speakers are more likely to cross the legs at the level of the ankles (35.71%) compared to listeners (13.46%). Listeners are more likely to maintain other crossing forms such as crossing at the level of the knees (30.77%) compared to speakers (25%), passing ankles over thigh (11.54%) compared to speakers (3.57%) or crossed leg

(21.15%) compared to speakers (10.71%). The results from leg-to-leg distance and crossing-leg forms are thus consistent (for example, listeners cross their legs at the levels of the knees and thus have knees together).

Table 16: Leg-crossing forms

	Ankle over thigh	At knees	At ankles	Crossed legg	No crossing
speaker	3,57%	25,00%	**35,71%**	10,71%	25,00%
listener	**11,54%**	**30,77%**	13,46%	**21,15%**	23,08%

In term of the orientation of the legs (as shown in Figure 20 **b**), speakers are more likely to direct legs to the side (27.12%) compared to listeners (12.31%). Listeners are likely to have more diversified postures in which the legs are directed to the front (44.62% compared to speakers 37.29%).

Table 17: Leg orientation

	Out	Side	Front	Inward
speaker	6,78%	**27,12%**	37,29%	28,81%
listener	10,77%	12,31%	**44,62%**	32,31%

In terms of right leg swivel (as shown in Figure 20 **c**), speakers are more likely to maintain their feet inside the knee (62.96%) compared to listeners (49.12%). Listeners are more likely to maintain their feet in line with the knee at 90 ° (43.86%) compared to speakers (35.19%), and have the feet outside the knee (7.02%), compared to speakers (1.85%).

Table 18: Leg swivel

	Feet outside the knee	Feet inside the knee	Feet in line with the knee (default)
speaker	1.85%	**62.96%**	35.19%
listener	**7.02%**	49.12%	**43.86%**

In terms of leg distance (as shown in Figure 20 **d**), speakers are more likely to maintain postures in which legs stand backward (19. 61%) compared to listeners (12%). In contrast, listeners are more likely to maintain legs bent at 90 ° (64%) compared to speakers (58.82%).

	Feet behind the knee	Feet in front of the knee	Feet in line with the knee
speaker	**19,61%**	21,57%	58,82%
listener	12,00%	24,00%	**64,00%**

The main differences between speakers and listeners in terms of the legs postures at the level of the lower body are illustrated in the figure below.

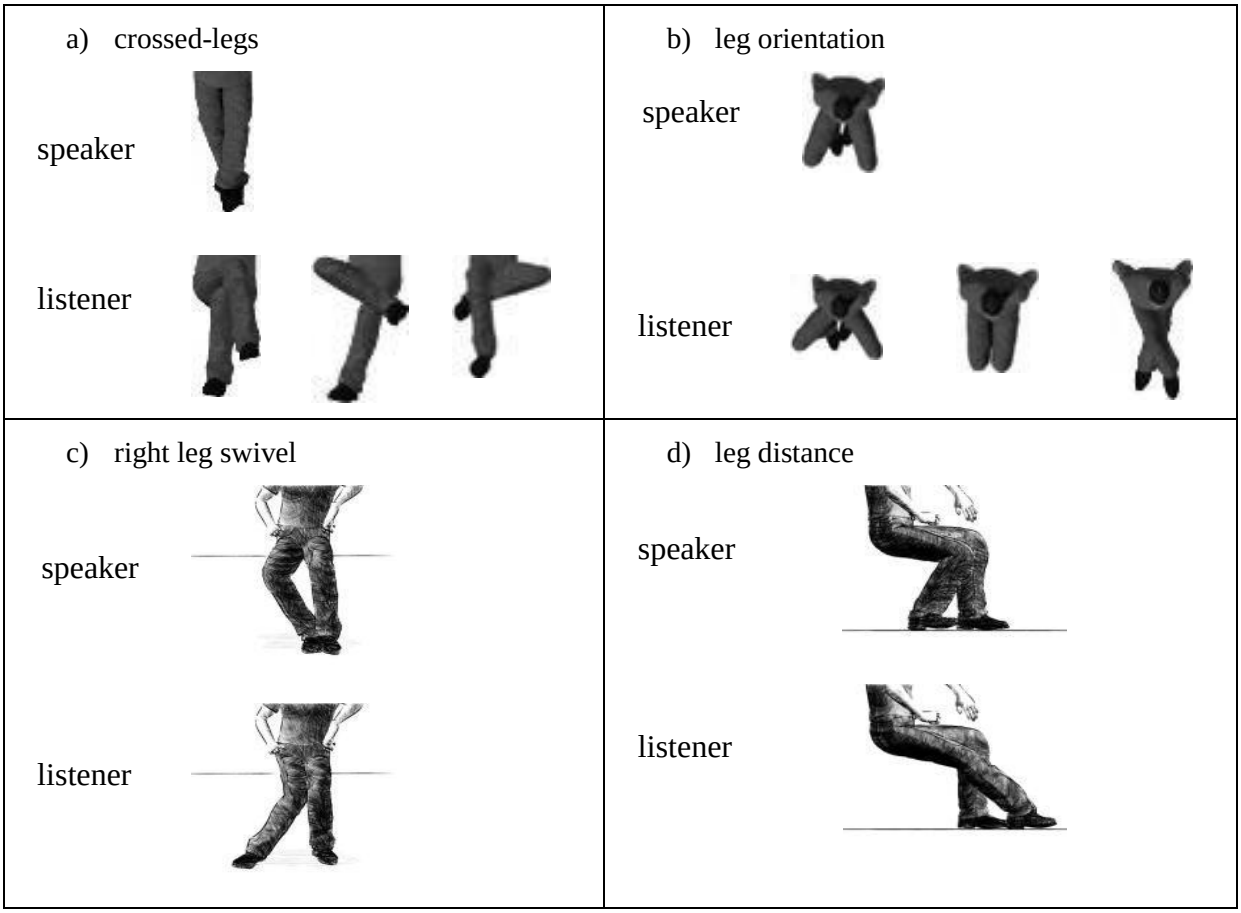


Figure 20 : Main differences between speakers and listeners in terms of leg postures

II.3.2.4 Discussion

We aimed to investigate the relationships between the body postures and the speaker / listener role. Association analyses were performed on a set of significant contingency tables extracted from Anvil involving 23 attributes of postures and the feature of the speaker / listener role. 21 out of 23 features were found to be statistically associated with that of the speaker / listener role.

Figure 19 and Figure 20 illustrate the main findings of the qualitative analysis on the association between the speaker/listener role and the different attributes of body postures. These results suggest different body posture recurrent patterns according to the speaking role that the subject plays.

Previous studies in linguistics analyzed how body postures define units within interactions. For example, some studies observed some synchronization between the locution cluster (paragraph) and body postures (Scheflen 1964; Kendon 1972; Kendon 1980). Kendon also observed a consistency between the use of the arm and the use of body postures in synchronization with locution (paragraph). This observation involves the highest level in the hierarchy of the alignment of gesture and intonation: “consistent arm use and body posture” and “locution cluster” (Table 19).

Instead of considering the locution as in the study by Kendon, we investigated the speaker / listener role that correspond to the speaking turn of the interlocutors in a dyadic conversation. Turn-taking involve a higher level than locution clusters in terms of social communicative functions.

Loehr (Loehr 2004) observed that people, when they expressed speech disfluency, performed a sewing-machine movement (foot on the floor, knee at a right angle and moving within the vertical plane) at the level of the leg. Even if this study focused on the dynamics (body movement) of the leg postures, which is beyond the scope of our work, we share the same perspectives with respect to the relation between speaking behaviors and leg postures.

Table 19: Kendon's Alignment of Gestural and Intonational Hierarchies (Loehr 2004)

Gesture	Intonation	Notes
Stroke	Stressed Syllable	
Gestural Phrase (preparation + stroke + retraction)	Tone Unit ("completed intonation tune")	A "tone unit" would probably correspond to an intonational phrase
Gestural Unit (from rest to rest)	Locution (sentence)	
Consistent Head Movement	Locution Group (common intonational feature)	A "locution group" would consist of sentences with "parallel" intonation
Consistent Arm Use and Body Posture	Locution Cluster (paragraph)	A "locution cluster" would probably correspond to a "discourse segment"

II.3.3 Data preparation for the analyses about convergence

We exported the annotation files for both postures and the speaker / listener role from Anvil. Four video-based annotations files of 60 minutes involved the AB and CM subjects: we obtained 4 data files per subjects. In total, 89682 frames (25 frames per second) were extracted from the annotations. Twenty-six features (called attributes in Anvil) of postural descriptions and 1 feature of the speaker / listener role were associated to each frame. The feature of the speaker/listener role was dummied as "1" for speaker and "2" for listener. Values of each of the 26 features were also dummied varying from "1" to "5". For each data file, we had a resulting matrix of 89682*27 values. The characteristics of the features were already detailed previously, and can be briefly summarized here:

- arm posture * 5 features
- shoulder posture * 1 feature
- trunk posture * 1 feature
- leg posture * 7 features
- speaker/listener *1 feature

II.3.4 Algorithm for estimating convergence

We defined an algorithm to process the matrix obtained in the previous section in order to estimate the convergence between the two subjects.

$DataPersonA(t, f)$ is a matrix that contains in each slot the value of the feature f at the time-frame t for the subject A.

Each row of the matrix concerns the same time-frame t.

Each column of the matrix contains 26 values (one value per feature).

We defined the following algorithm to estimate the convergence for each feature, when the person A is speaker and person B is listener. Thus, we considered separately the convergence when subject A is speaking while B is listening and the convergence when subject A is listening and subject B is speaking.

Let $ConvergenceCount := [c_1 \quad \cdots \quad c_{26}]$ be a vector of 26 elements.

Each element of the vector is the number of convergence that is observed for the corresponding feature. These values are set to 0 at the start of the algorithm.

Output: $ConvergenceCount$ = CONVERGENCE_COUNTER ($DataPersonA, DataPersonB$)

FOR t : $each\ time\ frame\ of\ DataPersonA$

FOR f : $each\ feature\ of\ DataPersonA$

IF both $DataPersonA(t, f)$ AND $DataPersonB(t, f)$ contain annotated data

IF A is speaker and B is listener at the time t

IF ($DataPersonB([t + 1, t + D], f) = DataPersonA(t, f)$ for D > 25) [3]

$ConvergenceCount(f) \mathrel{+}= 1$

END

END

END

END

END

For the values of t where the B is the speaker and A is the listener, the convergence count is computed by calling the same algorithm with CONVERGENCE_COUNTER ($DataPersonB, DataPersonA$).

[3] A posture is defined as any change that has been maintained within at least one second (25 frames = 1s).

II.3.5 Data analysis

We implemented this algorithm using Matlab and applied it to the postural annotations of two subjects in a one-hour dyadic conversation (4 videos AB-CM: 0-15, 15-30, 30-45, 45-60). We had a total of 602 postural convergences in terms of different forms and body parts. Table 20 presents the number of postural convergence for each body part.

Table 20: Numbers of postural convergence in 4 videos of the CID corpus (for a total of one hour)

	left body	right body	whole body	Percentage (%)
Arms Swivel	**36**	**37**	**73**	**22%**
Arms Radial Orientation	**36**	**33**	**69**	**20%**
Forearm Hand Orientation	**30**	**36**	**66**	**19%**
Arms Height	**29**	**31**	**60**	**18%**
Arms Distance	**25**	**26**	**51**	**15%**
Arms Touch	**9**	**9**	**18**	**5%**
Arms RadialZ	**1**	**1**	**2**	**1%**
TOTAL ARMS	**166**	**173**	**339**	**100%**
Legs Distance	**31**	**28**	**59**	**23%**
Legs Swivel	**28**	**30**	**58**	**22%**
Legs Height	**15**	**32**	**47**	**18%**
Leg To Leg Distance	**16**	**19**	**35**	**13%**
Crossed Legs	**11**	**21**	**32**	**12%**
Legs Radial Orientation	**20**	**12**	**32**	**12%**
TOTAL LEGS	**121**	**142**	**263**	**100%**
total	**287**	**315**	**602**	**/**

II.3.5.1 Individual differences

We also compared postural convergences between the two subjects. The two subjects were both right-handed. The results showed that participant CM displayed more often postural convergence than participant AB for most of postural features, except for the LegsRadialOrientation. This suggested that the subject CM has a higher tendency to imitate the behavior of her interlocutor than the reverse.

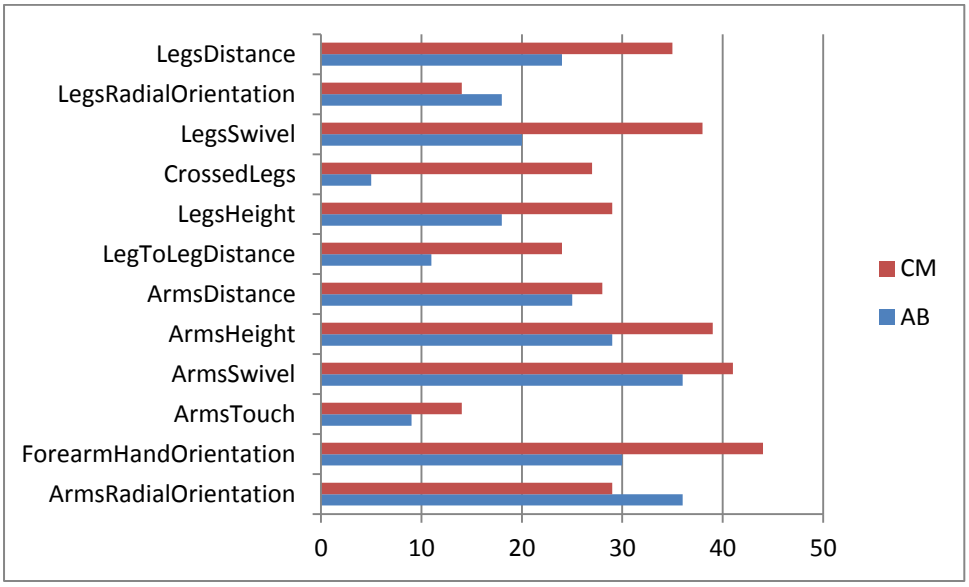


Figure 21: Number of postural convergence for AB when she is listening, and CM when she is listening, during a one-hour dyadic conversation

II.3.5.2 Upper body and lower body

We analyzed the differences of postural convergence between the upper body (shoulder, trunk and arms) and the lower body (legs).

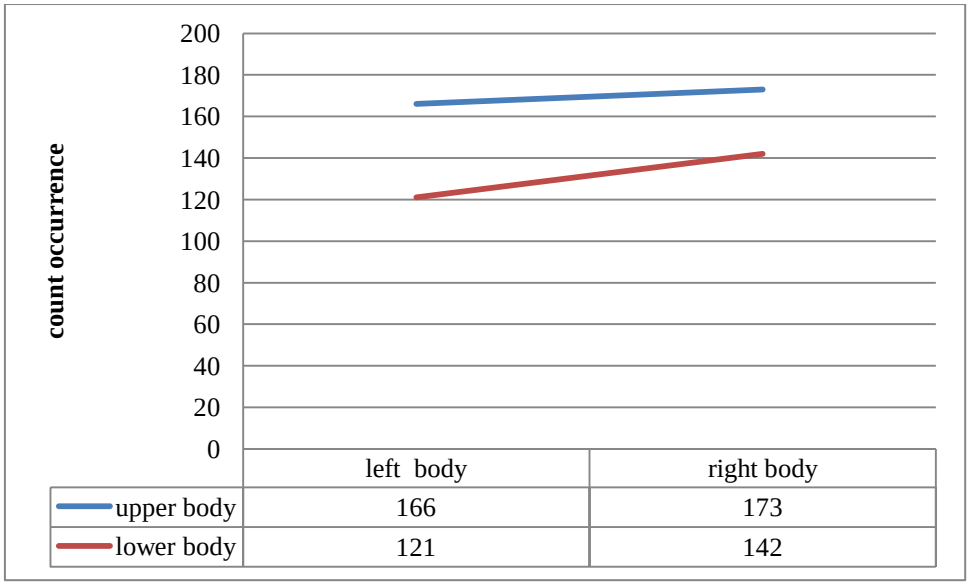
Concerning the upper body postures, we observed that postural convergence occurred mostly in terms of arm swivel (22%), arm radial orientation (20%), forearm / hand orientation (19%) arm height (18%) and arm distance (15%).

Concerning the lower body postures, we observed that postural convergence occurred mostly in terms of leg distance (23%), leg swivel (22%) and leg height (18%). Crossed legs (12%) and leg to leg distance (13%) are seldom observed during postural convergence. The spatial configurations of the lower leg of the listener adapt frequently to those of the speaker in the horizontal and vertical plans (leg distance and leg swivel).

More convergences were observed for the upper body (339) than for the lower body (263).

II.3.5.3 Right body vs. Left body

Figure 22 presents the number of estimated postural convergences according to right/left body and the upper/lower body. We observe that the frequency of postural convergence for the right body is higher than for the left body.



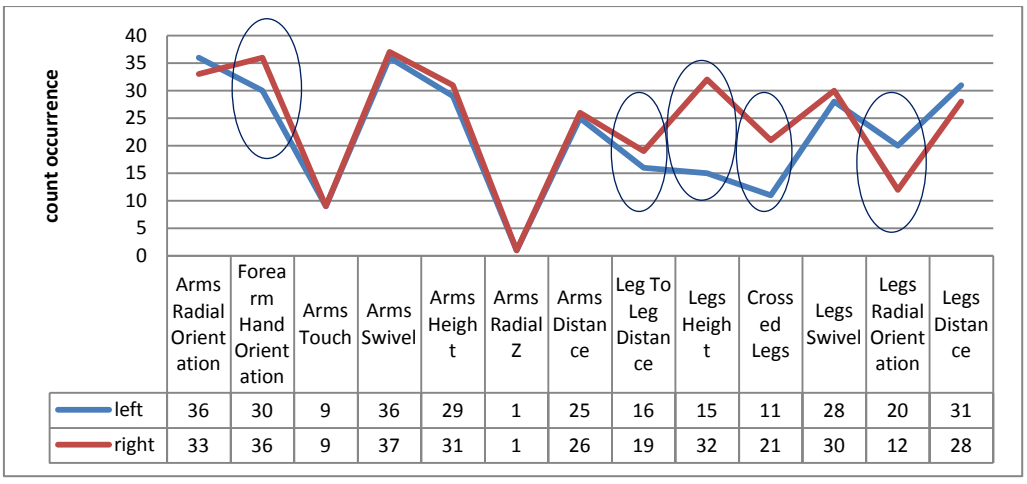
	left body	right body
upper body	166	173
lower body	121	142

Figure 22: Number of convergences observed according to left/right and lower/upper body

Figure 23 presents the results as a function of different features and the left/right body.

We found that 1) the difference between the left-upper and the right-upper body was due to the feature of forearm/arm orientation; 2) The difference between the left-lower and the right-lower body was due to the following features: leg-to-leg distance, leg height, crossed legs, legs radial orientation.

Interestingly, the features that are the mostly affected by postural convergence (arm swivel, arm radial orientation, leg distance and leg swivel) do not contribute to the difference between the right and the left body.



	Arms Radial Orientation	Forearm Hand Orientation	Arms Touch	Arms Swivel	Arms Height	Arms Radial Z	Arms Distance	Leg To Leg Distance	Legs Height	Crossed Legs	Legs Swivel	Legs Radial Orientation	Legs Distance
left	36	30	9	36	29	1	25	16	15	11	28	20	31
right	33	36	9	37	31	1	26	19	32	21	30	12	28

Figure 23: Number of postural convergence in 4 videos of the CID corpus (for a total of one hour) as a function of left/right body

II.3.5.4 Summary

We observed that postural convergence mainly occurs in terms of arm swivel, arm radial orientation, leg distance and leg swivel (Table 21).

Back to the example given at the beginning of this section (Figure 16): during this illustration of postural convergence, the listener adapted her posture to the posture of the speaker. Subtle differences exist between these two postures in terms of arm height ('above head' for the speaker and 'head' for the listener), but it shows the same configurations in terms of arm swivel ('raised') and radial orientation ('out').

This illustrative example is in line with our findings about the significant attributes of the postural convergence. These main findings are summarized in Table 21.

Table 21: Main attributes where postural convergence is observed

	attributes
Upper body	arm swivel
	arm radial orientation
Lower body	leg distance
	leg swivel

II.3.6 Discussion

We investigated to what extent the listener would imitate the posture of the speaker in a dyadic conversation. We coded the postures of the speaker and the listener roles in a dyadic conversation corpus. The symbolic coding system that we defined includes four different body parts (shoulder, trunk, arm and leg) with 26 different features that describe spatial configurations of these body parts.

We counted the number of postural convergence in terms of each of these 26 features. The results showed that the listener mostly adapted his postures to those of the speaker by pivoting arm up/down, pivoting arm front/back, moving shin along the horizontal axis and the saggital axis.

The effect of postural convergence is not affected by the impact of the right/left body. The features that are discriminative to the effect of the right/left body are different to those involving the effect of postural convergence. We also found that the features concerning touching (i.e. arm touch, crossed legs, leg-to-leg distance) are relatively independent from the postural convergence. These effects might result from other factors such as individual preferences of using body parts (handedness), the orientation of body in dyadic conversation and affect.

Our results show some differences between the two interlocutors. Individual differences might affect the production of postural convergence. However, the effect of individual differences was not investigated further in this study, because it only involved one pair of participants.

This study need to be extended following the perspective that is suggested in the study by (Mol, Krahmer et al. 2011). These authors observed that the perspective of the confederate's gestures influence the hand shape of participants' gestures. In their study, the authors argued that repeating other's meaningful gesture forms is not simply a case of direct mimicry of form. Instead, perceiving a gesture gives rise to the convergence of semantic representations. They asked a couple of confederate and a subject to describe route directions by using different hand shapes (finger vs. hand) and perspectives (map vs. route). The perspective of the confederate's gestures did have an impact on the hand shape of the participants' gestures. This result was explained by the understanding of the meaning of confederate's gestures (map) that leads the participant to make vertical gestures as on a map using the finger more frequently.

Future analyses should be conducted to relate the observed postural convergences to semantic convergences. These analyses will require the annotation of other linguistic levels and other means to estimate convergence.

II.4 Thesis contribution: the EXPO-standing scheme for coding standing leg postures

The EXPO-sitting scheme described in the previous section focused on the postures observed during seated interaction. We were willing to study standing postures in order to have a more global and more generic scheme for representing postures in different contexts.

II.4.1 A coding scheme for standing leg postures

Coding leg posture depends on the fact that the recorded person is sitting or standing. We did not find similar coding scheme for standing leg postures. Most existing corpora consider only sitting postural expressions.

We named our new scheme for coding leg postures in a standing context the **EXPO-standing scheme**.

Inspired by Bull's scheme (Bull 1987), we selected same three salient elements for coding a posture of a leg that are already in the EXPO-sitting scheme: the knee, the feet, and the overall leg posture.

Coding leg postures in a standing situation raises different requirements than coding leg postures in a sitting context. We described the overall leg posture in terms of its role (support, light or normal), the position of the leg compared to a vertical axis (front, normal, back), and the shape of the legs (cross, apart, closed and normal). The right leg and the left leg are coded using the same scheme but along with separate tracks.

We considered three forms of the knee: slightly flexed, flexed or normal form.

We describe the foot according to its orientation (open, closed, normal) and its shape (flat on floor, tiptoe, toes and heel).

Table 22: The EXPO-standing coding scheme

$$LEGS = \left\{ \begin{array}{c} legrole = normal, support, light, upbeat \\ legshape = normal, crossed, closed, apart \\ legposition = normal, forward, backward \\ kneeshape = normal, slightly\ flexed, flexed \\ footorientation = normal, opened, closed \\ footshape = flat\ on\ floor, heel, tiptoe, toes \end{array} \right\}$$

Table 23: Illustration of values for the leg-role attribute for the left leg

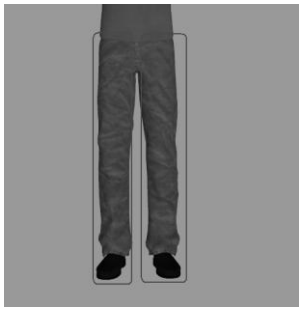
Values	Support	Normal	Light	Upbeat
The attribute describes how the center of mass shifts above the standing supporting leg.				

Table 24: Illustration of values for the leg-position attribute for the right leg

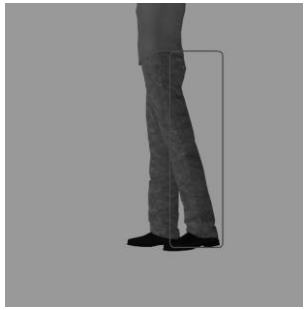
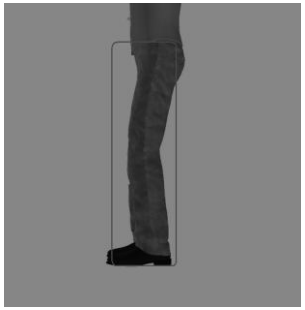
Values	Forward	Backward	Normal
The attribute describes if the leg is drawn back or stretched out			

Table 25: Illustration of values for the leg-shape attribute for the right leg

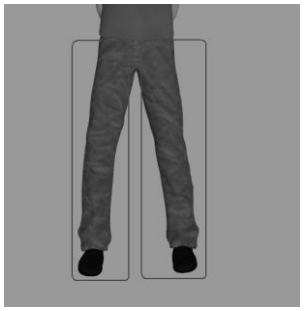
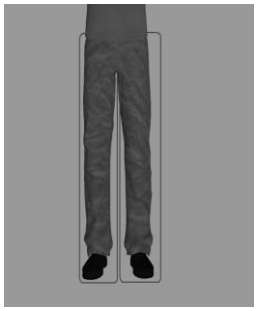
Values	Crossed	Apart	Closed	Normal
The attribute describes the postures of one leg relatively to the posture of the other leg				

Table 26: Illustration of values for the knee-shape attribute for the left leg

Values	Normal	Slightly flexed	Flexed
The attribute describes how the knee is bent.			

Table 27: Illustration of values for the foot-orientation attribute for the right leg

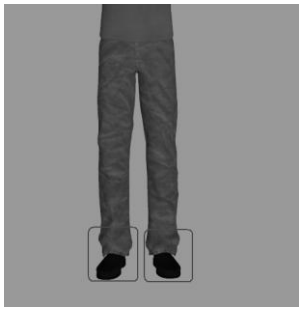
Values	Normal	Opened	Closed
The attribute describes the orientation of the foot. It is defined relatively to a position in which the foot is pointed straight ahead.			

Table 28: Illustration of values for the foot-orientation attribute for the left leg

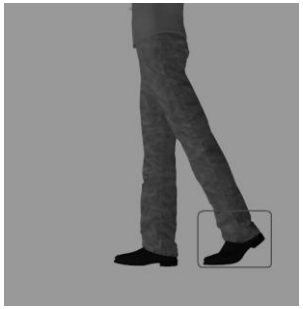
Values	Flat on floor	Tiptoe	Toes	Heel
The attribute describes how the toes of the foot are positioned with respect to the floor.				

Table 29: An example of annotation of leg postures using the EXPO-standing coding scheme

	Right leg	Left leg
Role of leg	**Light**	**Support**
Position of leg	**Forward**	**Normal**
Form of leg	**Crossed**	**Normal**
Knee	**Slightly flexed**	**Normal**
Orientation of feet	**Opened**	**Normal**
Form of feet	**Tiptoe**	**Normal**



Table 30: A second example of annotation of leg postures using the EXPO-standing coding scheme

	Right leg	Left leg
Role of leg	**Support**	**Upbeat**
Position of leg	**Normal**	**Normal**
Form of leg	**Normal**	**Normal**
Knee	**Normal**	**Flexed**
Orientation of feet	**Normal**	**Normal**
Form of feet	**Normal**	**Normal**



II.4.2 Collection of video corpus featuring standing postures: the story-telling corpus ContAct

Studying leg postures in standing positions required the collection of another corpus that the CID corpus.

GV-LEX is a project funded by the French National Agency for Research (ANR) involving Aldebaran Robotics (coordinator), Acapela, Telecom Paris Tech and LIMSI-CNRS. It aimed to make a humanoid robot and a virtual character capable of reading texts for several minutes without boring the listener thanks to expressive gestures and speech.

To inform the design of the gestures to be expressed by storytelling characters or robots, we collected a video corpus called ContAct (Conte Actés / Acted Stories). The corpus was collected at LIMSI and used by the partners in charge of gesture generation. We were also interested in having variations of the different ways to tell the same story. We did not find any available video corpus of that kind (e.g. full body expressions displayed by various persons telling a story) so we decided to collect our own.

Six actors from an amateur theatre troupe were recruited to be recorded. Each actor told the same story twice so that we would get possible variations by the same actor. They were asked to tell a story using their whole body. They were filmed from two different angles (front and profile) to get a good visibility of the whole body posture. The story “Three small pieces of

night" written by a professional story teller was selected due to its wide range of emotional content (positive and negative emotions).

The collected corpus is composed of 11 videos involving six different actors: two videos for five actors (the sixth actor declined to collect a second recording). The average duration of each clip is 7.5 minutes. Table 31 provides a detailed description of the corpus. The figure below provides sample frames of this corpus.

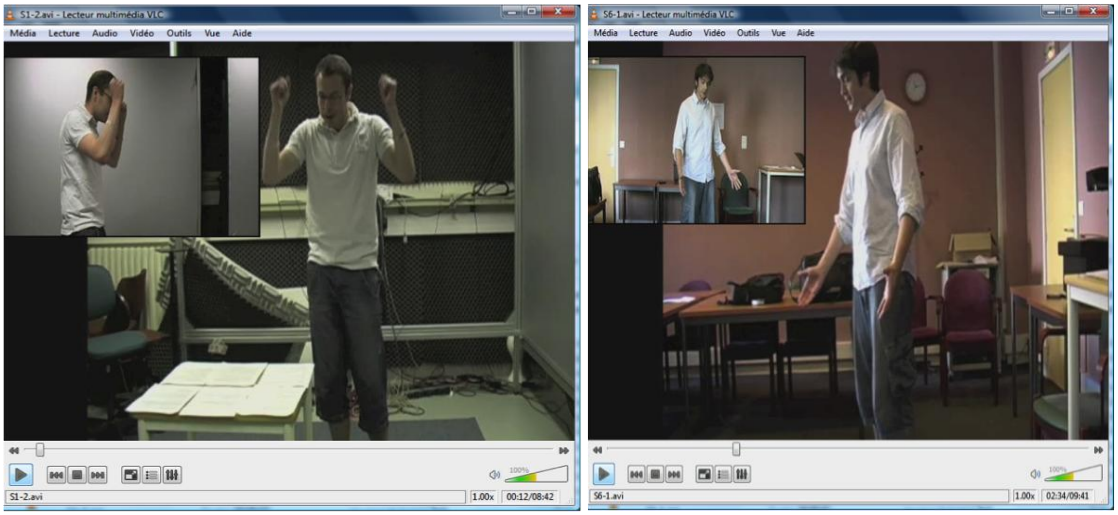


Figure 24: Frames of the ContAct corpus

Table 31: Corpus descriptions

Clips	Min.	Sec.	TOTAL SEC
S1-1	8	20	500
S1-2	8	42	522
S2-1	6	18	378
S2-2	6	15	375
S3-1	6	33	393
S3-2	7	12	432
S4-1	6	28	388
S4-2	6	58	418
S5-1	7	41	461
S5-2	8	11	491
S6-1	9	41	581
TOTAL	82,3		4939 (82'19")
Average	7,5		449 (7'29")

II.4.3 Annotations leg postures in the ContAct corpus

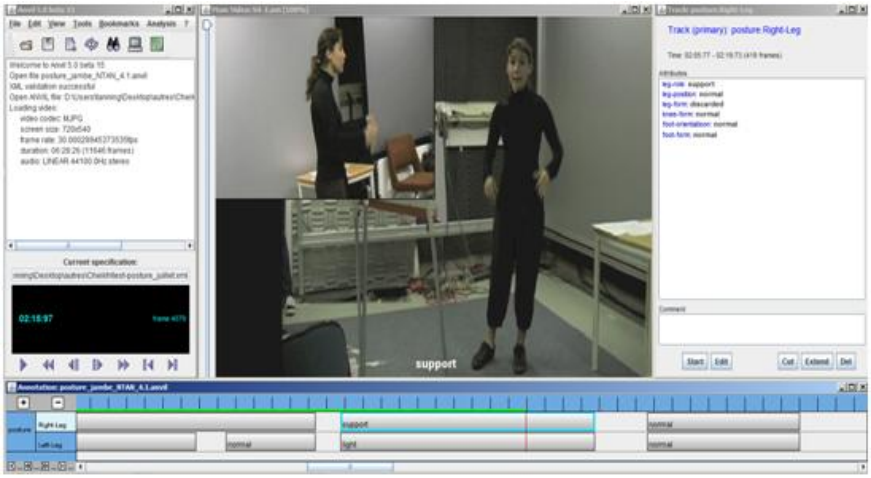


Figure 25: An example of leg posture annotations in the story-telling corpus ContAct

We performed a manual annotation of the five videos that were acted by three participants (two participants acted two sessions). We selected these clips due to the good visibility on the legs. 365 segments were annotated with respect to leg postures using the EXPO-standing scheme, (Table 32).

Table 32: A summary of leg posture annotations based on the 5 recording clips

Clips	right leg	left leg	total
s2-2	25	24	49
s3-1	19	19	38
s3-2	45	43	88
s4-1	23	21	44
s4-2	73	73	146
Total	185	180	365

II.4.4 Validation of the EXPO-standing coding scheme

In order to valid our EXPO-standing coding scheme, a second coder annotated the postures of the right leg in one video (S3-1).

As shown in the table above, the first coder had annotated 19 segments for this track in this video. The second coder annotated 18 segments.

A strong agreement was observed for the EXPO-standing coding scheme with a kappa value of 0.802 for the annotated leg. Three attributes received a sustained intercoder agreement (when k value is above .80): the leg position (0.804), the knee shape (0.863) and the foot shape (0.806).Three attributes received a high level of intercoder agreement (when k value is between .60 and .80): respectively, the leg-status (0.703), the leg shape (0.757), and the foot orientation (0.776).

Table 33: Kappa values based on the annotations of leg postures by two coders

Attribute	Kappa (k)
leg role	0.703
leg position	0.847
leg shape	0.757
knee shape	0.863
foot orientation	0.776
foot shape	0.863
AVERAGE	0.802

II.4.5 Clustering of leg postures

II.4.5.1 Objective

Literature on non verbal studies suggested individual differences in terms of nonverbal behaviors (Knapp and Hall 2006). These differences have been the focus of some studies, for example gesture individual profiles based on video corpora (Kipp 2001). Yet, few studies in the Multimodal Corpora field explore the notion of individual postural profiles. The ContAct corpus is relevant for such exploratory studies since several actors tell the same story.

The approach that we selected was to investigate the most common legs postures that subjects take up (nonetheless of the time duration of that posture), which have been defined as **posture types**: a typical posture generalized based on the posture observations.

The EXPO-standing coding scheme leads to annotating separately the two legs based on a set of 6 attributes (leg role, leg position, leg shape, knee shape, foot orientation, and foot shape). In this study, a leg posture is thus represented as a vector of 12 features.

We use the same data analysis process as for the EXPO sitting coding scheme.

II.4.5.2 Data extraction

We exported a total of 60243 frames from Anvil based on the leg posture annotations of the five clips. The detailed information is shown in Table 34.

We then retrieved the 60243 instances according to the criteria that have been described previously:

- Exclude the frames in which there is no annotation in any track;
- Exclude the successive frames that contain the same annotations.

These extractions lead to 140 significant posture instances. The matrix in the extracted data set contained a matrix of 140–by-12. The 12 features were composed of 6 attributes with respect to the two legs according the EXPO-standing coding scheme (right-leg-role, left-leg-role, right-leg-position, left-leg-position, right-leg-shape, left-leg-shape, right-knee-shape, left-leg-shape, right-foot-orientation, left-foot-orientation, right-foot-shape, left-foot-shape).

Table 34: Number of instances of leg postures before and after the extraction

Clips	Number of instances in the original data set	Number of instances in the extracted data set
s2-2	11271	20
s3-1	11808	15
s3-2	12969	35
s4-1	11645	22
s4-2	12550	48
TOTAL	60243	140

II.4.5.3 Data analysis

We applied a Hierarchic Clustering on the extracted data. This method clusters different posture combinations into similar categories. We created a Hierarchical cluster tree (Figure 26) using the Euclidean distance metric and Ward's linkage algorithm.

We found three gross *posture types*. Each U-shape line represents the distance between the two posture combinations. The branches are collapsed according to distance. Each point on the X axis represents one instance (out of the 140 postures).

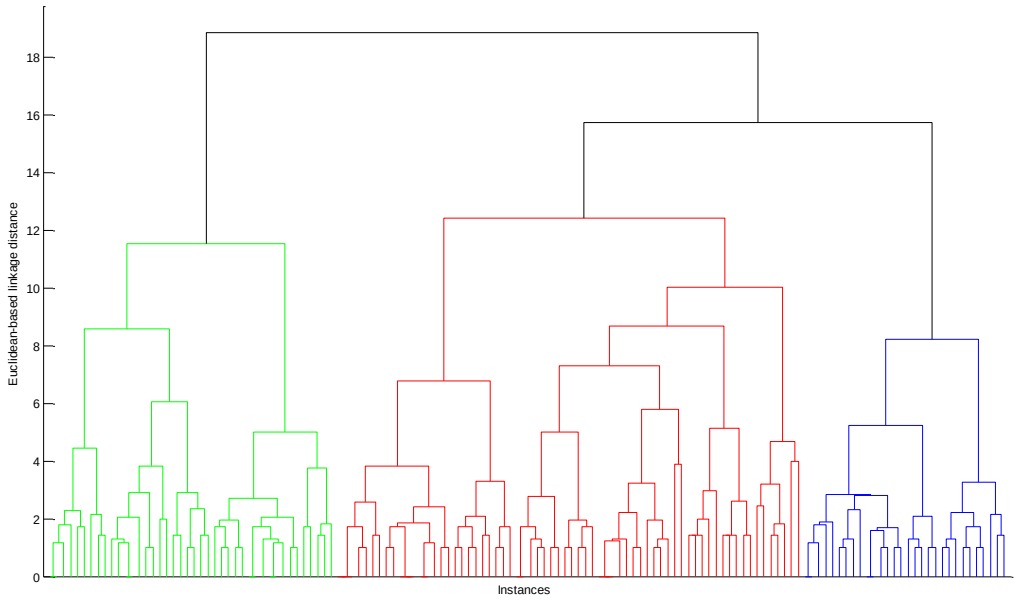
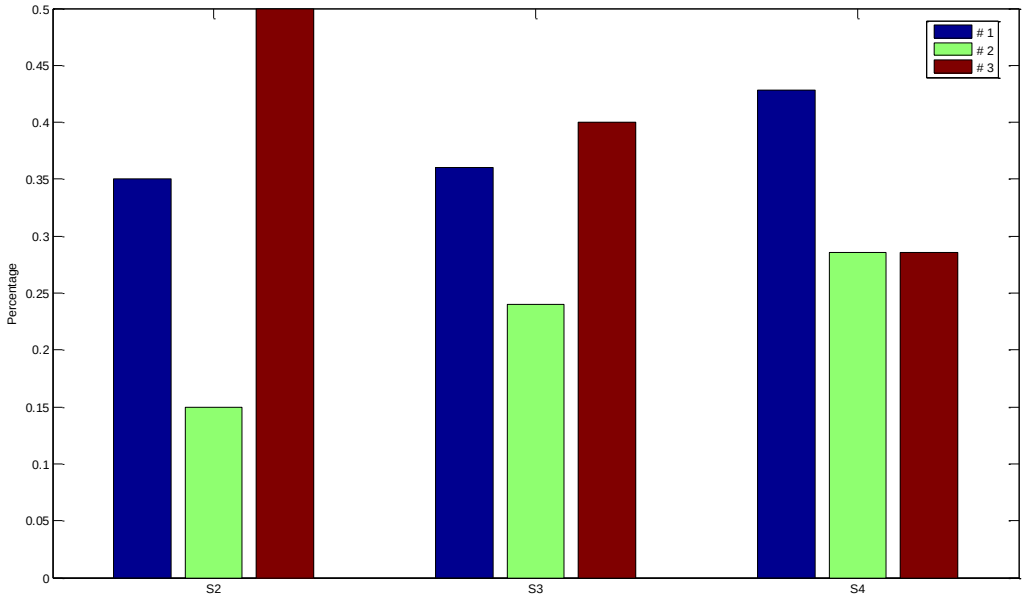


Figure 26: Dentogram of the clusters based on the 140 extracted posture instances from the annotations of 5 clips (different colors represent different posture types)

The results of this clustering analysis marked each posture instances with a posture type code ("1" means posture type #1, "2" stands for posture type #2, and "3" posture type #3). As shown in Table 35, subject S2 contributed much more to posture type #3 (50%) than to posture type #2 (15%). Subject S4 contributed more to posture type #1 (42.86%) than the two other posture types (28.57% and 28.57%). Overall, the three subjects took up less frequently the posture type #2 than the two other posture types.

Table 35: Percentage of the three posture types (posture type #1, posture type #2 and posture type #3) expressed by subjects (S2, S3, and S4)

	#1		#2		#3		TOTAL	
Subjects	Nb. of posture instances		Nb. of posture instances		Nb. of posture instances		Nb. of posture instances	
S2	7	35.00%	3	15.00%	10	50.00%	20	100%
S3	9	36.00%	6	24.00%	10	40.00%	25	100%
S4	15	42.86%	10	28.57%	10	28.57%	35	100%
TOTAL	55		35		50		140	



The 140 posture instances were then plotted in a 2-dimensional space using a Principal Component Analysis (PCA) .In the figure below, the red points refer to posture type #1, which represents 48.57% of the 140 posture instances. The green points refer to posture type #2, which represents 30% of the 140 posture instances. The blue points refer to posture type #3, which represents 21.42% of the 140 postural instances. The X and Y axes represent the first two principal components that explained roughly 47.8% of the total variance.

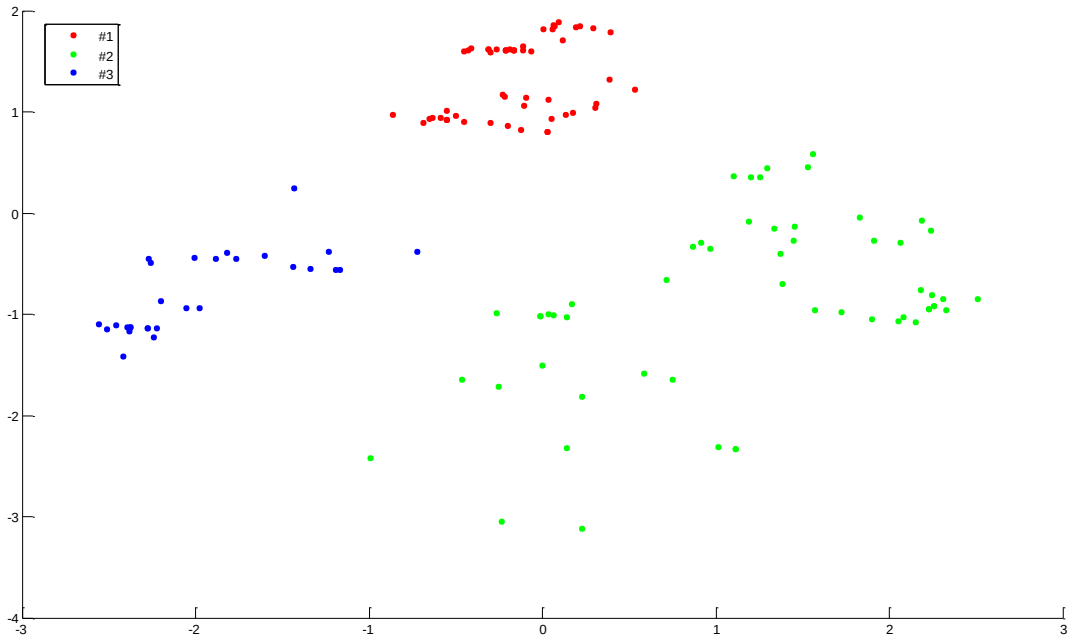


Figure 27: Mapping the 140 extracted posture instances on the 2-dimensional principal component space

We selected three points from each of the clusters to visually illustrate the posture types. Each point refers to a particular posture for which we could find the matching frame in the video clips. As shown in Table 36, three illustrations were retrieved per subject.

Table 36: Descriptions of the 3 posture types based on the 140 extracted posture instances

Posture type	#1	#2	#3	TOTAL
S2				
S3				
S4				
Number of posture instances	68	42	30	140
Percentage	48.57%	30%	21.42%	100%

We then investigated the contributions of each of the 12 leg posture features. In order to visualize the magnitude and sign of the contributions of each feature to the components, we computed the contribution scores of each feature. The scores are shown in Table 37. For example, the Principal Component #1 had the largest positive coefficients corresponding to the feature *right-leg-shape* (0.7087), and the largest negative coefficient corresponding to the feature *left-leg-shape* (-0.6074). The results of the PCA analysis showed that the first four principal components already contributed to 72% of the total variance involving the features:

- the right leg shape,
- the left leg shape,
- the right leg role,
- the left leg role
- and the right leg position.

Table 37: PCA scores: transforming the original data into the space of the principal components (the values in bold are those of highest absolute value per component).

ORGINAL FEATURES	COMPONENT 1	COMPONENT 2	COMPONENT 3	COMPONENT 4
right-leg-role	0.0907	-0.261	**0.6374**	-0.1127
right-leg-position	0.1786	0.1148	0.248	**0.8377**
right-leg-shape	**0.7087**	**-0.6275**	-0.2019	-0.0496
right-knee-shape	0.0722	-0.0477	0.0539	0.0492
right-foot-orientation	0.1512	-0.0128	0.0157	0.2438
right-leg-shape	0.0494	0.0632	0.1232	0.2639
left-leg-role	-0.1645	-0.2198	**0.6434**	-0.2016
left-leg-position	-0.1354	-0.0232	0.0416	0.1158
left-leg-shape	**-0.6074**	**-0.683**	-0.2178	0.3025
left-knee-shape	-0.0487	0.0548	0.0187	-0.0362
left-foot-orientation	-0.0996	0.0131	-0.0939	-0.0689
left-leg-shape	-0.0217	0.0004	0.0264	0.0024

II.4.6 Conclusion

In section II.4, we achieved the following goals. First, we investigated whether there are particular and common postural cues at the level of legs during monologues. We computed a clustering algorithm on the 140 extracted postures instances (based on the 20643 original instances) based on 5 clips of a story-telling corpus. Three gross posture types were found. This result showed that speakers used a small set of shapes and orientations of postures at the level of lower body over a large variety of forms of legs that they might have. This finding is in line with the results of section II.2.3 about full body postures.

Second, we validated the EXPO standing coding scheme using a Principal Component Analysis on the 140 extracted postures. The results showed that five out of the 12 features should be considered as the significant contributors to the principal components. These features mainly involve the shape, the role and the position of the leg. Whereas the features with respect to the knee shape, the foot orientation and the foot shape were meant to be eliminated from the descriptions of the leg postures during a monologue in our case of study.

II.5 **Conclusion**

In this chapter we described our contributions to the symbolic description of static postures and their convergence during a spontaneous conversation.

First, the EXPO scheme was defined to provide a formal description of the visual appearance of the whole body posture. It employs the single static posture act as its basic unit. The EXPO scheme is descriptive and does not contain any features in relation to the semantics or the assumed social meaning of the body forms. The agreement between the human coders was measured by computing Cohen's kappa on some manual annotations, thereby providing a validation of parts of our scheme.

Second, the EXPO scheme was applied in two different settings: sitting postures during dyadic conversations, and standing postures during story-telling.

The EXPO-sitting scheme aims to symbolically describe the postural configurations that are observed at the level of the arms, the shoulders, the trunk and the legs by two sitting interlocutors having a spontaneous conversation in French. The EXPO-standing scheme aims to describe the postural configurations that are observed at the level of the legs during storytelling monologues. These complementary schemes are expected to provide a comprehensive description of the visual appearance of sitting and standing postures.

One of the major challenges for the formal description of body postures is to find an appropriate level of description. Our approach combines an exhaustive symbolic description of the whole body and a principal component analysis that allows eliminating the redundancy features without losing essential information. We annotated postures using our EXPO-sitting and EXPO-standing schemes. We then applied a simple clustering algorithm on these annotations. The data was extracted from the original annotation data set, so that it only involves the posture frequency with which a given posture occurred and was taken up (but not the total length of time for which a posture is maintained). We found that only a small set of posture combinations are commonly used by each of our subjects in the recorded context. Such results might be of interest in future studies for the design of the repertoire of bodily expressions that is required for virtual characters.

We also highlighted the role of space in developing the posture coding system. We distinguished the space of gestures (*gesture space*) and the space of postural convergence (*shared space*). We developed a scheme for coding the gesture space based on the propositions by McNeill. We observed that coding gesture space using videos remains limited for describing the variations along the saggittal plane.

We were also interested in the psycholinguistic phenomenon of convergence that occurs in a *shared space*. We developed a method to compute postural convergence during dyadic conversation based on the symbolic annotations of postures. The results showed that the listener did tend to adapt her postures to those of the speaker. We confirmed the assumption that postures, like speech, may allow for the convergence of representations of meaning across interlocutors. Further studies are needed to investigate how postural convergence occurs in relation to convergence in other modalities such as speech convergence or facial expression.

We used a video corpus because it might enable subjects to gesture more freely and have a more natural dyadic conversation that when being equipped with motion capture sensors. This advantage enables naturalistic studies on the encoding and decoding of human postures. Nevertheless, complementary studies need to be conducted using motion capture data to validate and complement the results that we found using manual annotations of videos.

Furthermore, individual differences should be explored in terms of individual postural behaviors, in line with similar studies conducted on individual gesture profiles (Kipp 2001).

This chapter contributed to the understandings of body posture in human interactions without considering the underlying semantics. In the next two chapters, we present their applications in two different contexts of human-computer interactions: affective interaction and ambient interaction.

III. BODILY EXPRESSIONS IN VIRTUAL CHARACTERS: APPLICATION TO AFFECTIVE INTERACTION

"*Our natural way of thinking is that the mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression. My thesis on the contrary, is that the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion.*"

William James, What is an emotion? (1884)

III.1 **Related Work**

III.1.1 Emotions

According to (Fehr and Russell 1984), everyone knows what an emotion until asked to give a definition is. The nature of emotions has been continuously debated. In 1884, William James published a paper entitled "What is emotion?". He wondered if the emotion is the cause of the behavioral response or its consequence. For example, when faced with a bear, do we flee because we are afraid, or are we afraid because we flee.

Among the several definitions of emotions, we selected the following one for this thesis. An emotion can be seen as "an episode of interrelated, synchronized changes in five components in response to an event of major significance to the organism " (Scherer 2000). These five components are: the cognitive processing (function of evaluation of the objects and of the events), the physiological changes (function of system regulation), the action tendencies (function of preparation and direction of action), the motor expression (function of communication of reaction and behavioral intention) and the subjective feeling (function of monitoring of internal state and organism-environment interaction).

Approaches and theories of emotion are surveyed in several recent papers (Scherer 2010; Gross and Barrett 2011) that enable to compare them (Figure 28).

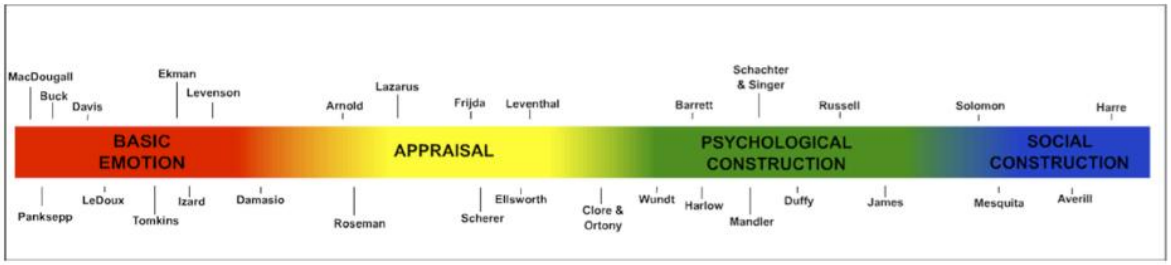


Figure 1. Perspectives on emotion can be loosely arranged along a continuum. We have populated this continuum with representative theorists/researchers drawn from the field of psychology. We distinguish four "zones": (1) basic emotion, in red, e.g., MacDougall (1908/1921), Panksepp (1998), Buck (1999), Davis (1992), LeDoux (2000), Tomkins (1962, 1963), Ekman (1972), Izard (1993), Levenson (1994), and Damasio (1999); (2) appraisal, in yellow, e.g., Arnold (1960a, 1960b), Roseman (1991), Lazarus (1991), Frijda (1986), Scherer (1984), Smith and Ellsworth (1985), Leventhal (1984), and Clore and Ortony (2008); (3) psychological construction, in green, e.g., Wundt (1897/1998), Barrett (2009), Harlow and Stagner (1933), Mandler (1975), Schachter and Singer (1962), Duffy (1941); Russell (2003), and James (1884); (4) social construction, in blue, e.g., Solomon (2003), Mesquita (2010), Averill (1980), and Harré (1986). Given space constraints, as well as the goals of this article, we have limited ourselves to a subset of the many theorists/researchers who might have been included on this continuum (e.g., those who only study one aspect of emotion were not included in this figure).

Figure 28: A continuum proposed for representing the theories of emotions (Gross and Barrett 2011)

Another point of view consists in representing the different models of emotions using a table representing how these models involve a) the different components of emotions (cognitive, physiological, expressive, motivational / action tendency, feeling), and b) the different phases of cognitive evaluation (Scherer 2010) (Figure 29).

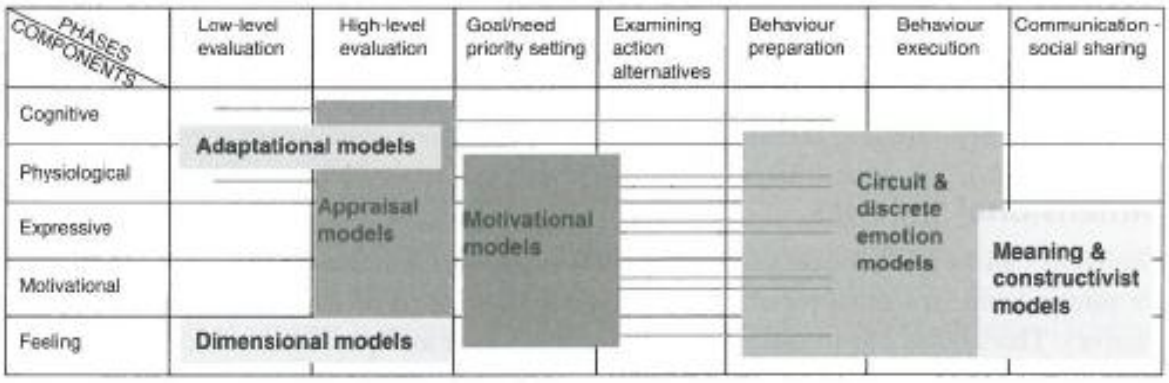


Figure 29: Models of emotion represented according to the components of emotions (lines) and the phases of cognitive evaluation (columns)(Scherer 2010).

Figure 30 summarizes different computational models and their relation with psychological theories (Marsella et al. 2010).

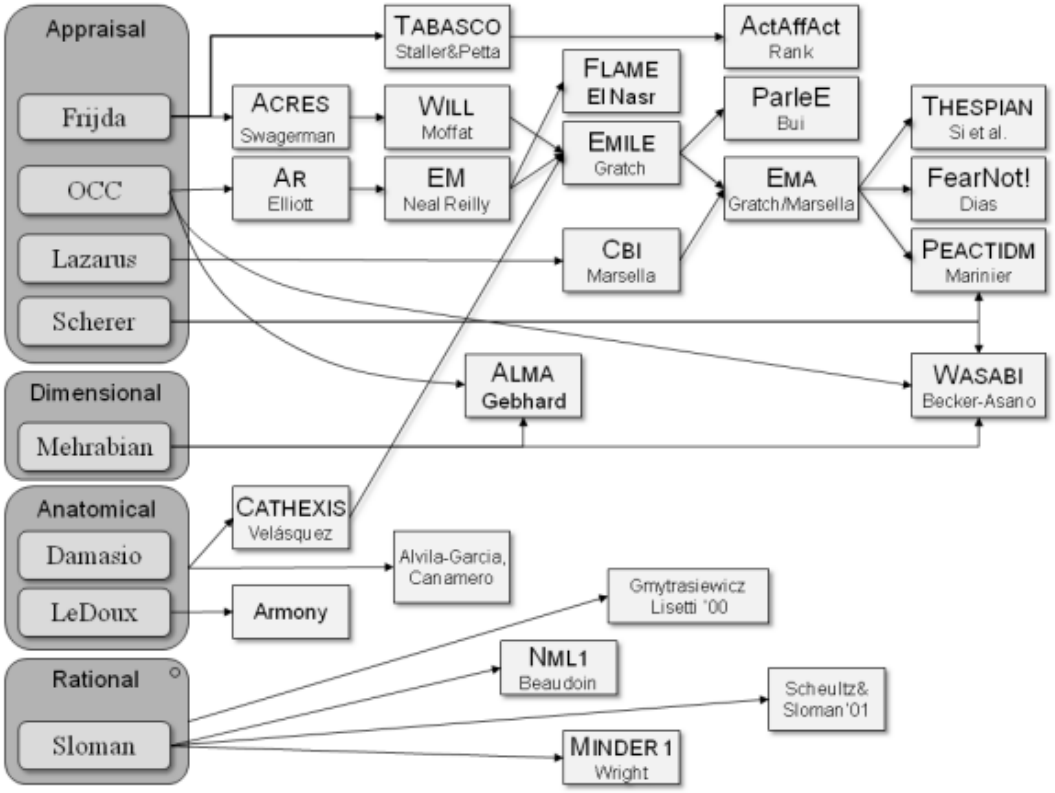


Figure 30: Computational models of emotions and their relations with the underlying psychological theories (left) (Marsella, Gratch et al. 2010).

In the next section, we focus on one of the components of emotion: the action tendency. As we will explain, this component sounds promising for conducting original studies about postural expressions. Postures were observed to convey subtle nonverbal behaviors and hence are good candidates for possibly conveying subtle signs of behaviors preparation.

Yet, action tendencies received less attention from researchers in affective computing than the other components of emotion. Some researchers study action tendencies but in terms of facial expressions (Tcherkassof 2008; Meillon, Tcherkassof et al. 2010).

Thus, following the grid proposed by Figure 29, this chapter of the thesis focuses on the behavior preparation and execution phases and the motivational component of emotions.

We will first describe this specific component of emotion, and then survey the different studies about bodily expressions of emotion.

III.1.2 Action Tendencies

According to Frijda, *action tendencies* are "states of readiness to execute a given kind of action, and thus an action tendency is defined by its end result aimed at or achieved" (Frijda 1987). Action tendencies play a role in the preparation and the direction of action at the motivation level. On the one hand, they consist of a readiness to execute action: they involve

the activation of a class of possible responses selected out of a human's response repertoire. On the other hand, they consist of "readiness to achieve or maintain a given kind of relationship with the environment" (Scherer 2000). Thus, they involve orientation toward a present or forthcoming state. They can be seen as plans or programs to achieve such ends, which are put in a state of readiness. They aim at achieving changes to the actual situation. Action tendencies are thus "readiness to engage in or disengage from interaction with some goal object in some particular fashion". They are changes in action readiness (activation or inhibition states) to engage in interaction with the environment (Frijda 1987).

Different action tendencies correspond to different emotions. Some emotion categories might be better represented by appraisals (jealousy, surprise, hope), and others by action tendencies (disgust, despair, anxiety, anger) (Frijda, Kuipers et al. 1989).

Since there are links between action tendencies and emotion categories, there are also possible links between classes of action tendencies and some features common to several emotion categories, for example the valence.

Darwin suggested in his famous "the Expressions of Emotion in Man and Animal" that emotional expressions do have two functional aspects: preparation of adaptation behavior and regularization of social interactions (Darwin 1872). Based on the emotion expressions of others, we predict their specific action tendencies, and use them as guides to act in the interactional events. Lazarus claimed that if emotions do have adaptive functions, they should bring about action tendencies that allow the individual to cope with the eliciting event (Lazarus 1991). The action tendencies in turn should produce differentiated patterns of responses and expressive signals in the different modalities.

III.1.3 Understanding the emotions expressed by the body

Bodily expression of emotion is addressed by multiple disciplines and research areas. In this section, we briefly survey some studies in Psychology, Neurosciences, Arts. We also discuss some research and techniques for the animation of virtual characters.

III.1.3.1 Psychological studies

Posture was observed to be a relevant modality to express various emotion categories. As we described previously, bodily expressions involve several channels, such as hand gesture (Kendon 1980; McNeill 1992), posture (Bull 1987), and body movements (Ekman 1965). Whereas expressions of different emotion categories in facial expressions, gaze, and gestures were explored in several studies, postural expressions of emotion and other components of emotions (e.g. action tendencies) received less attention.

Darwin suggested that specific body movements are associated with each emotional state (Darwin 1872). For example, joy can be expressed by various purposeless movements such as: jumping, dancing for joy, and clapping the hands. The expression of Sadness might be motionless and passive, having the head hang over a contracted chest. Pride is often expressed with an erected head and body. Shame can be expressed by turning away the body or the face.

Ekman suggested that body actions might provide information about the intensity of the felt emotion (Ekman 1964). Ekman put forward the view that the face is more important than the body in conveying emotion categories. He suggested that judgments based on the face lead to a higher recognition accuracy and a higher agreement among viewers when considering the emotion category (Ekman 1964), whereas the body would play a role in conveying the

intensity of the emotion (Ekman 1965). Ekman and Friesen suggested that static postures are more likely to convey gross affect (e.g. liking), whereas movements of the body are more likely to convey specific emotions (Ekman and Friesen 1967).

Bull (Bull 1987) reported several studies about Bodily expressions. In one of these studies, he found for example that interest can be communicated by leaning forward and drawing back the legs. He defined a coding system for body postures and movements associated with specific emotions categories.

Wallbott observed discriminative features of emotions both in static body postures and in the movement quality (Wallbott 1998). The corpus consisted of 224 video recordings. The elicited emotions were derived from (Scherer 2010), including elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride and boredom. The bodily expressions of these emotions by actors were analyzed. Discriminative features were found for several emotion categories. For examples, sadness was related to collapsed upper body and low movement dynamics. Anger was related to lateralized movements of hands with arms stretched out frontally. Pride was related to crossed arms in front of the chest and head directed towards the back. Elated joy was related to shoulders up, arms stretched out frontally or upward, illustrator gestures, high movement activity, expansive movement and high movement dynamics.

Coulson (Coulson 2004) investigated whether changes in the viewpoint of postures would affect the accuracy of emotion recognition. He compared the perception of static postures expressing six emotions viewed from three different angles. Sixty-one participants viewed 528 images of postures (thirty-two postures for anger, disgust, happiness, sadness and twenty-four postures for fear and surprise, each posture was then rendered from 3 viewing angles). Correlations across the three viewpoints for each emotion category showed that, for anger, happiness and surprise, the attributions were more likely when the participants perceived postures from the front. For fear, the attributions were more likely when the postures were seen from behind and from the side. For sadness, the attributions were more likely when the postures were perceived from the front and from the side (Table 38).

Table 38: Prediction of five emotions with the best viewpoint and the corresponding anatomical features (Coulson 2004)

Emotions	Viewpoint	Head bend	Chest bend	Abdomen / twist	Shoulder adduct/abduct	Shoulder swing	Elbow bend	Mass center movement
Anger	more likely when seen from the front	backwards	absence of backward	no	arms raised forwards and upwards			either forwards or backwards
Fear	less likely when seen from the front	backwards	no	no	forearms are raised		no	either forwards or backwards
Happiness	more likely for Front view	backwards	no forwards		arms raised at the shoulder level		straight	no
Sadness	less likely when viewed from behind	Forwards	forwards	no	arms at the side of the trunk			no
Surprise	more likely for Front view	backwards	backwards	any degree	arms raised		straight	no

Scherer (Scherer, Schorr et al. 2001) considered the causal sequences of bodily changes in his Component Process Model of emotion (Table 39). He predicted a series of bodily posture shifts following a process of multi-level sequential checking. For example, a successful coping of an individual with a stimulus event might lead to “agonistic hand/arm movements, an erect posture, and a lean forward of the trunk”. As previously described, appraisal is based

on five groups of SECs (Stimulus Evaluation Checks) to generate emotional states. First, the individual checks if the stimulus is novel (*novelty check*). Second, the individual checks if the stimulus is pleasant or negative (*intrinsic pleasantness check*). Third, the individual checks if the stimulus is relevant for a goal or need (*goal significance check*). Four, the individual checks if the stimulus is under control (*coping potential check*). This check separately checks the cause of existence of a stimulus (causality sub-check), gauges the coping potential that is available to an organism (control sub-check), measures the energy of the organism that can be mobilized in order to change or to avoid negative consequences through fight or flight (power sub-check), and gauges to which degree an organism can accept a new outcome by adaptation, gauges to which degree the organism can accept a new outcome by adaptation (adjusting sub-check). Five, the individual checks if the stimulus agrees with *norms and self*.

Table 39: Predictions of body changes following major SEC outcomes (Scherer 2010)

Stimulus Evaluation checks		Changes on the Body	
Novelty Check	**Check if the stimulus is novel**	**Novel**	**Not novel**
		Interruption of ongoing instrumental actions, raising head, straightening posture	NONE
Intrinsic Pleasantness Check	**Check if the stimulus is pleasant or negative**	**Pleasant**	**Unpleasant**
		Centripetal hand and arm movements, expanding postures, approach locomotion	Centrifugal hand and arm movements, hands covering orifices, shrinking postures, avoidance locomotion
Goal/Need Conduciveness Check	**Check if the stimulus is relevant for a goal or need**	**Relevant and discrepant**	**Relevant and consistent**
		/	Strong tonus, task-dependant instrumental actions
		Comfort and rest positions	/
Coping Potential Check	**Check if the stimulus is under control**	**No control**	**Control and high power**
		Few movements, slumped posture	Agonistic hand/arm movements, erect posture, body lean forward, approach locomotion
			(High power) Protective hand/arm movements, fast locomotion or freezing
Norm / Self Compatibility Check	**Check if the stimulus agrees with social norms and internalized norms**	/	/

III.1.3.2 Neuroscience studies

The role of body postures has been reinforced over facial expressions in cases of incongruent affective displays (de Gelder, Snyder et al. 2004).

Stienen and de Gelder (de Gelder and Van den Stock 2010) used the blindsight approach to investigate whether human viewers could detect affect information independently of subjective visual awareness. The subjects listened to a short story before visualizing a set of images. A total of 224 photosets were presented to 23 subjects, who had to detect the emotion expression and to indicate their confidence. The stimuli contain a target image (selected from eight images of a fearful bodily expression); a distractor image (selected from eight images of expressions of combing the hairs) and a pattern mask (completely covering the area of the body). The face was covered with an opaque oval patch. The photoset consists of 14 timing conditions: the 12 different value settings of time duration between the target/distractor image and the pattern mask; two other control conditions include target-only and mask-only displays. A trial then includes: 1) 500 ms of initial delay, 2) 33 ms of the target display, 3) an interval between the display of the target/distractor image and that of the pattern mask (-50, -33, -17, 0, 17, 33, 50, 67, 83, 100, 117, 133 ms), When the value is (-33, -17, 0, 17), the mask and the target/distractor image is overlapped 4) 50 ms displaying the pattern mask 5) 17 milliseconds or 767 ms of gray screen. The results showed that the fearful bodily expressions are being processed independently of subjective visual awareness. This indicated that briefly seen, but also consciously unseen bodily stimuli may induce an emotional state and trigger adaptive actions in the observer, potentially involving sub cortical visual structures.

Other researchers (Kret, Pichon et al. 2011) investigated the neural areas dedicated to the processing of threatening bodily expressions (fear and anger). These authors used magnetic resonance imaging methods to study the sensitivity of several specific neural areas. The results showed that:

1) the amygdala was more active for facial than for bodily expressions ; this is in line with previous finding by Adolphs, Tranel and Damasio (Adolphs, Tranel et al. 2003) who showed that a patient that has a damaged amygdala is able to recognize emotions from a dynamic facial expression but not from a static one

2) the extra striate body area (EBA) was more active for threatening versus neutral expressions and for bodies than faces,

3) the superior temporal sulcus (STS) was active for dynamic threatening body expressions, and sensitive to affective information conveyed by the body stimulus ; previous studies reported that the superior temporal sulcus focuses on goal-directed actions and configuration and kinematic information from body movement.

These studies in Neurosciences underlined that dynamic bodily expressions are different from static ones. As the same as suggested Phenomenology, our nature lies in movement and bodily states are not static (Pascal and Gleason 1966).

As we saw in this section, studies in Neurosciences investigated the extent to which neural areas predict emotions in the perception of the human body. Bodily expressions are analyzed independently of facial expressions in these studies. These studies focus on several negative emotions (i.e. fear and angry). The assumption is that humans are best at detecting potentially harmful information (i.e. negative emotional expressions) from bodily expressions. Yet, the perception of other emotions has received little attention from the neuroscientists.

III.1.3.3 Coding bodily expressions in Arts

Laban Movement Analysis (LMA) was initially developed by Rudolf Laban to interpret, describe, visualize and notate the more subtle characteristics about the way a human movement is done with respect to its inner intention (Newlove 1993). According to Laban's conception of human movement, human movement is a fluid, dynamic transiency of simultaneous changes in spatial positioning, body activation and energy usage. Analyzing human body movement requires two levels: 1) what the movement is made of ; even if the human being moves always to satisfy a particular need, some basic elements of physical action are common to all human motions ; and 2) how movements are put together: laws of sequencing and the rhythmic patterns provide a governing order that prevents the movement from being chaotic (Zhao 2001). The Laban Movement Analysis includes four dimensions for describing human movement (Table 38).

Table 40: Laban Movement Analysis (Body, Effort, Shape and Space)

Dimensions	Descriptions
Body	The body system describes which body parts are moving (initiation of movement starting from specific body parts), which parts are connected (connection of different body parts to each other), which parts are influenced by others (sequencing of movement between parts of the body), and general statements about body organization (patterns of body organization and connectivity).
Effort	The effort system consists of four bipolar factors: space (indirect vs. direct), weight (light vs. strong), time (sustained vs. sudden), flow (free vs. bound). Effort Actions incude 8 combinations of the first three factors ("space", "weight" and "time"): for example float, punch, glide, slash, dab, wring, flick and press. The factor "flow" describes the continuousness of motions. Any movement may contain a combination of these effort factors.
Shape	The Shape system has several subcategories: 1) the shape forms describe the static shapes that the body takes; 2) the modes of shape change describe the way the body is interacting with the environment, 3) the shape qualities describe the way the body is changing toward some point in space (e.g. opening, closing, rising, sinking, spreading, enclosing, advancing and retreating), 4) the shape flow support describes the way the torso can change in shape to support movements of the rest of the body.
Space	The Space system involves motions in connection with the environment, with spatial patterns, pathways, and lines of spatial tension.

Delsarte's system of expressions is based on systematic observations of the human body. The system influenced acting and modern dance by linking meaning and motion. Each gesture is expressive of an internal meaning (emotions and traits). This internal meaning includes the meaning of different body parts and different directions of movements. For example, the *excentric* movement (away from the body center) is related to the exterior world. The *concentric* movements (towards to the body center) is related to the interior. According to the

Delsarte's system, nine posture combinations per body parts can provide meaning (Nixon, Pasquier et al. 2010).

III.1.4 Techniques for encoding emotions in bodily expressions

III.1.4.1 The illusion of life and digital animation techniques

The encoding of bodily expressions in animated characters was studied in the traditional animation techniques and in the digital animation. In early animations there was no movement in the figures but a simple progression across the paper. The animators had no knowledge about the relationships between forms with flows of action from one drawing to another. They used the same cartoon figure in a new position on a next piece of paper.

Walt Disney suggested a set of methods for relating drawings to each other, which became afterwards the fundamental twelve principles of animation (Thomas and Johnston 1995): squash and stretch, anticipation, staging, straight ahead action and pose to pose, follow through and overlapping action, slow in and slow out, arcs, secondary action, timing, exaggeration, solid drawing, and appeal. Walt Disney considered that timing can determine whether the character is excited, relaxed or nervous. He believed that the *inbetween* drawing not only relate one drawing to another, but also gives a new meaning to the action.

Another famous animator Walt Stanchfield suggested using at least three drawings to study and portray each body action: 1) Preparation – telling the audience something is going to happen, 2) Anticipation – gathering the forces to carry through with the action, and 3) Action – carrying out of the intended action (Stanchfield and Hahn 2009).

The above studies built ground in developing computational generations of behaviors in virtual character by modeling every single key frame. The process was extended to 3D computer animation, for which the spatial juxtaposition of objects in a scene can be defined by key frames and the computer can interpolate the in-between frames to control the timing. The problem is that linear interpolation generally leads to unrealistic animations.

III.1.4.2 Multimodal corpora and automatic analysis methods

Digital corpora and machine learning techniques are sometimes combined to reveal the critical features for the decoding of emotional body expressions. Researchers recorded digital corpora to provide detailed information about the expression of emotions through different modalities in humans. Automated methods are then developed to predict emotions using multimodal nonverbal cues including facial expressions, hand movements, postures, skin conductance and mouse pressure data (Castellano, Leite et al. 2010).

The GEMEP corpus (Geneva Multimodal Emotion Portrayal database) contains video recordings of 18 acted emotion portrayals by 10 professional actors (Bänziger, Pirker et al. 2006). The major challenge of the corpus is situated at the descriptive level of the nonverbal behaviors. It contains not only audio-video recordings of emotional expressions but also descriptive data such as facial, gestural and acoustic features as well as the judgments about the accuracy and the genuineness of perceived emotions.

(D'Mello and Graesser 2006) developed a multimodal affect detector that combines conversational cues, gross body language and facial features. 28 participants interacted with

an intelligent tutoring system in a conversational dialogue for a session of 32 minutes. The authors used linear discriminant analyses to compare fixed and spontaneous experiences of five emotion categories: boredom, engagement/flow, confusion, frustration, delight and neutral. The results showed that posture was redundant with two other channels in both the fixed and the spontaneous contexts. The model combining facial features with contextual cues was the best emotion detection strategy. However, posture alone was able to detect boredom and engagement, but not confusion and neutral.

The AffectMe project collected a library of bodily expressions of emotions using a VICON motion-capture system (Bianchi-Berthouze and Kleinsmith 2003). 13 subjects acted anger, fear, happiness and sadness. Their movements were recorded at 32 points of the body. 111 affective postures were collected and presented to five subjects who had to judge these postures according to emotion categories and affective dimensions (valence, arousal, potency and avoidance). The results showed that some bodily features provide information about some affective dimensions. For example, openness of the body seemed to be important to arousal dimension of affective. In addition, the researchers (Kleinsmith, Bianchi-Berthouze et al. 2010) collected a second non-acted corpus of subtle body expressions. Ten subjects played a Nintendo Wii game for 20 minutes. The subjects were not aware of the purpose of the study so that they would be spontaneous. For each movement, three temporal segments were chosen to build the training set of postures: a starting posture, a posture occurring in the middle of the movement, and the posture at the apex of the movement. Eight subjects annotated the collected postures. The resulting training set was composed of 32 defeated, 33 triumphant and 33 neutral postures.

In the following tables, we summarize several corpora of postures that are relevant to our research goals (e.g. studying postures during conversations or expressions of emotions). For each of these corpora, we indicate information about behaviors and functions, elicitation methods, data size, coded features, research domains, descriptions of posture, and the analyses that were conducted.

Table 41: Studies and corpora of bodily expressions of emotions (1)

Database	Emotional content	Emotion collection methods	Data size	Coded features	Areas	Posture definition	Classifiers and analyses	
(Wallbott 1998)	Cold anger, hot anger, elated joy, happiness, disgust, contempt, sadness, despair, fear, terror, shame, interest, pride, boredom	Scenario-based elicitation with given utterances	224 postures: 14 emotions * 2 sentences * 2 scenarios * 2 genders of actors * 2 takes each (12 professional actors subjects, 6 male, 6 female)	Upper body (away from camera, collapsed), shoulders (up, backward, forward), head (downward, backward, turned sideways, bent sideways), arms (lateralized hand/arm movements, stretched out frontal, stretched out sideways, crossed in front of chest, crossed in front of belly, before belly, stemmed to hips), hands (fists, opening/closing, back of hand sideways, self-manipulator, illustrator, emblem, index finger pointing), movement quality judgment (movement activity, expansiveness/spatial extension, movement dynamics/energy/power)	Social psychology	Body posture is distinguished from body movements. The posture categories consider both functional (emblem, illustrator) and anatomical aspects.	**Initialization:** Manual (12 expert coders with intercoder agreement test)	ANOVA
(Picard 2003)	Interested, Bored, Thinking, Seated, Relaxed, Defensive, Confident	Sitting pressure distribution maps	400 postures: 150 samples * 6 postures (30 children subjects, 15 female, 15 male)	6 features: upright, leaning forward, leaning left, leaning right, leaning back, slumping	Human Computer Interaction	Posture is defined as a series of body movements rather than a series of static positions.	**Initialization:** EM (Expectation Maximization),	Neural network for static postures; HMM (Hidden Markov Models) for continuous postures
(Coulson 2004) 		Poser 4 figure animation	528 postures: 32 samples * 4 emotions (anger, disgust, happiness, sadness), 24 samples for 2 emotions (fear, surprise)	Head bent, chest bent, abdomen twist, shoulder swing, shoulder, adduct/abduct, elbow bent, weight transfer	Social sciences	3D presence changes in viewing angle; movement of the mass center, and six joint rotations (head bend, chest bend, abdomen twist, shoulder adduct/abduct, shoulder swing, and elbow bend)	Pre-defined **Initialization**	Multinomial Logistic Regression

Table 42: Studies and corpora of bodily expressions of emotions (2)

Database	Emotional content	Emotion collection methods	data size	low-level postures description	domain	Posture definition	Classifiers and analyses	
(Kleinsmith and Bianchi-Berthouze 2007) 	Anger, Fear, Happiness, Sadness, Affective Dimensions: Valence, Arousal, Potency, Avoidance	Elicited emotions with 32 markers and 8 cameras of motion capture	108 affective postures: 27 samples * 4 postures (13 adult subjects exposed)	24 features describing the 3-orthogonal planes and the lateral, frontal, vertical extension of the body, body torsion and the inclination of the head and shoulders	Posture Recognition	Posture is defined as any stance involving the hands and/or body that can convey emotions or feelings	**Initialization:** Unsupervised clustering algorithm, EM (Expectation Maximization)	MDA (Mixture Discriminate Analysis)
(Kleinsmith, Bianchi-Berthouze et al. 2010) 	Concentrated, frustrated, Defeated, triumphant	Elicited emotions with 32 markers and 8 cameras of motion capture (Gypsy 5) in a whole-body computer game scenario	103 non-acted postures including 32 defeated, 33 triumphant and 33 neutral postures	41 features describing 3D rotational information for each body joint (neck, collar, shoulders, elbows, wrists, torso, hips and knees)	Posture Recognition	Posture is defined as any stance involving the hands and/or body that can convey emotions or feelings	Multi-task learning (MTL)	
(Sanghvi, Castellano et al. 2010) 	Engagement	Interaction between children and a robot during a game Postures of users were videotaped from the front and from the side	44 recordings with a lateral view of the body which have been coded as “engaged with the robot”	Quantity of motion, posture leaning angle; curvature of the back; expansion or contraction of the upper body data normalization (e.g., whether the data was normalized or not), number of histogram bins, combination of features considered, combination of derivatives considered and number of frames used to compute the SMI in the QoM calculation.	Human Computer Interaction	A 7-second thin slice (body movement)	the ADTree and OneR classifiers	ADTree with the features of posture leaning angle and curvature of the back OneR with the features of quantity of motion and expansion or contraction of the upper body

III.1.4.3 Techniques for Postural Expression of Emotion by Interactive Virtual Characters

Several virtual character platforms consider how to express emotions using hand and head gestures, body postures and movement quality. For example, the Greta agent uses a model for expressive gesture based on several parameters of expressivity (Pelachaud 2005). Methods have also been proposed for generating expressive stance (Neff and Fiume 2006), idle movements (Egges, Papagiannakis et al. 2007) and for managing bodily expressions of interpersonnal attitudes (Gillies, Crabtree et al. 2006).

The Expressive Motion technique (Zhao 2001) consists in overlapping an original movement with a secondary expressive movement. The aim is to add expressiveness to animations by mapping high-level dimensions (e.g. emotions, personalities) to low-level parameters (i.e. the actual mathematical transformations and modifications of coordinates in space, joint angles, the actual number of frames used for a gesture, and the formulas of expressiveness). The use of secondary movements aims at contributing to the believability of the virtual characters. The combinational results of two levels of movement produce varied movements according to the virtual character's emotions, personalities or environment. Besides, other techniques are used to generate expressive movement of virtual characters, such as controlling behaviors and adding expressiveness to neutral motions. However, those techniques require expertise in modifying low-level configurations, they are difficult to use in generating the range of expressivity of human bodily expressions.

The EMOTE model was established based on the Effort and Shape systems of Laban Movement Analysis (Chi, Costa et al. 2000). It is based on the key-frame approach for generating animations. The considered parameters are independently set up and scaled according to the body parts. The coordination of those parameter settings results in the expressive behavior of the virtual characters. Effort parameters determine the wrist rotations and the number of frames between two key frames. Shape parameters link to arms and determine the space used by a gesture (i.e. distance of a key point towards the body centre) while considering the human constraints a priori. Shape parameters also link to torso with different values in vertical/horizontal/sagittal dimensions. For example, adjusting angles of neck, spine and clavicles makes torso turned sideward. Flourishes parameters are miscellaneous including two parameters of wrist bending (wrist bend multiplier and wrist extension magnitude) and four parameters of arm twisting (wrist twist magnitude, wrist frequency, elbow twist magnitude and elbow frequency).

This approach suggests a set of expressivity parameters of human bodily movement based on psychology literature. The animations and body movements are fully synthesized. Compared to the approach of EMOTE, the approach of ECA regard an intermediate level of behavior parameterization to map high-level qualitative communicative functions to low-level animation parameters. And particularly, it highlights the perceptual aspect of encoding movement. In other words, it does not focus on the underlying muscle activation patterns, but on the surface realization of movement. Hartmann et al. (2005) suggested a set of attributes to encode gesture expressivity: 1) overall activation: it quantifies movements during a conversational turn; 2) spatial extent: it describes amplitude of movements, 3) temporal extent: duration of the stroke phrase, 4) fluidity: smoothness and continuity of overall movement, 5) power: dynamic properties of the movements, 6) repetition: tendency to rhythmic repeats of specific movements.

The prominent work following this approach is that of Castellano (Castellano 2008) Castellano constructed a system, which acquires in real-time video input, extracts human movement characteristics and synthesizes the animation of an Embodied Conversational Agent (ECA). The system consists of two mapping levels: 1) mapping motion cues of extracted data (contraction index, velocity, acceleration, and fluidity) onto emotions (anger, joy and sadness, with different values: high/low/very high/medium); 2) mapping motion cues of extracted data (contraction index, velocity, acceleration, fluidity) onto the expressivity parameters (spatial extent, temporal extent, power and fluidity). An evaluation of perception was performed to investigate the extent to which people recognize emotions expressed by the agent using only body movement with the hidden face. This experiment contributed to revealing the impact of expressivity parameters in the communication and perception of emotions.

Table 43: A review of expressivity parameters in motor behavior and computational generation systems

Terms	References	Descriptions
Movement Duration / Temporal Extent	(Hartmann, Mancini et al. 2005)	Stroke phases are faster or slower.
Power		It refers to the dynamic properties of the movement: rest phases and continuity are affected without retraction time between two gestures.
Repetition		It refers to the tendency to rhythmic repeats of specific movements.
Spatial Extent		It refers to the amplitude of movements, the amount of space taken up by body, arms extend or contract towards the torso.
Overall activation		It refers to the quantity of movements.
Fluidity		It refers to the smoothness and the continuity of overall movement such as smooth/graceful vs. sudden/jerky.
Velocity (average)	(Doucet and Stelmack 1997)	Average velocity may be the key kinematic parameter in the initiation of movement, although this does not necessarily indicate that velocity is a parameter in response programming.
Amplitude of the Initial Impulse		The reaction time increases as the size of the target is decreased.
Movement Precision		The reaction time increases as the complexity increases.
Force Production		By achieving greater velocity the arm had produced greater acceleration and greater force at the same time. Consequently a greater amount of energy has been generated and utilized.

III.2 **Thesis Contribution: Designing static postural expressions of action tendencies**

As we saw in the previous section, several studies have explored how bodily expressions convey emotions in terms of categories, dimensions or appraisals. We did not find any study about the possible postural expressions of action tendencies, which is one of the components of emotions.

In this section, we explain the methodology we followed to design postural expressions of action tendencies for a virtual character. The next section will explain how these expressions were evaluated.

III.2.1 A Corpus of Bodily Expressions of Action Tendencies

We did not find much information in the literature about postural expressions of action tendencies.

In order to design the postural expressions of action tendencies of our virtual character, we used the PERMUTATION video corpus (Clavel and Martin 2009). This corpus contains 100 video samples of American television comedy-drama series for a total duration of 42 minutes. It contains samples of rich emotional interactions in a variety of social situations.

The corpus was presented to 200 participants, who attributed action tendencies based on the nonverbal behaviors of the characters. The protocol and the analyses are described in details in (Clavel and Martin 2009).



Figure 31: Frames from the PERMUTATION corpus.

We selected the five action tendencies that received the highest agreement among the subjects (Clavel and Martin 2009). We also selected the clips for which the upper or the whole body was visible and the action tendency well recognized. The selected action tendencies are listed in (Table 44).

Table 44: The five action tendencies selected for our study and the written description proposed by Frijda (Frijda, Kuipers et al. 1989)

Action tendencies	Description
Antagonistic	I wanted to oppose, to assault; hurt or insult.
Attending	I wanted to observe well, to understand, or I paid attention.
Disappear from view	I wanted to sink into the ground, to disappear from the Earth, not to be noticed by anyone.
Exuberant	I wanted to move, be Exuberant, sing, jump, and undertake things.
In command	I stood above the situation; I felt I was In command; I held the ropes.

III.2.2 Designing static postures of action tendencies

We used the MARC virtual characters platform for the design of our virtual agent and its postural expressions (Courgeon, Martin et al. 2008). We selected this platform because it enables to easily design static postures. It features several characters and 3D environment that are useful for situating the bodily expressions in a social context. It also enables to design dynamic expressions (described in the next section). Finally, the virtual characters can also be integrated in real-time interactive applications (as we will see in the next chapter of this thesis).

Using the MARC platform, we designed two postural expressions for each of the five selected action tendencies (Table 44). Our idea was to be able to compare the two postures designed for each action tendency and keep the best recognized for later use. Designing more than two postural expressions per action tendency would require a larger video corpus. One pair of neutral postures was also designed to control the experiment. The resulting 12 postural expressions are provided in Figure 32. The description of some of them using the EXPO coding scheme described in the previous chapter is provided in Table 46.

We situated these postural expressions in a 3D scene simulating a social interaction between two female characters standing face to face. Only the character facing the camera, displays emotional postures. The other character is visible from the back and is displayed only for helping the interpretation of postures thanks to the social interaction context. This second character stands in a neutral postural expression. Figure 32 presents the set of postural expressions designed for the selected set of action tendencies. The face was blurred to inhibit the influence of a facial expression on users' perception (even a neutral face might impact the perception of the postural expression (Clavel and Martin 2009).

The next section explains how we evaluated these images of postural expressions.

Table 45: Annotations of bodily expressions of action tendencies (for each body part, the most frequent annotations are listed)

Action tendencies (number of videos annotated with this action tendency)	Head	Trunk	Shoulder
Antagonistic (3 clips)	Tilt up-down, down	Forward	Forward and raised
Attending (2 clips)	Straightforward, tilt left down	Forward, sideways and raised	Forward
Disappear from view (3 clips)	Straightforward, shake, down	Forward, Backward	Raised and slightly forward
Exuberant (5 clips)	Straightforward	Forward	Forward
In command (1 clip)	Turn down sideways, right	Backward and raised	Backward and raised

Action tendencies	Postures	
Antagonistic		
Attending		
Disappear from view		
Exuberant		
In command		
Neutral		

Figure 32: Bodily expressions designed for the five selected action tendencies and for the Neutral

Table 46: Specification of some of the postures using the EXPO scheme

Antagonistic 	$\begin{bmatrix} \text{RIGHT ARM} & \begin{bmatrix} \text{ArmHeight} = waist \\ \text{ArmDistance} = touch \\ \text{ArmRadialOrientation} = behind \\ \text{ArmRadial Z} = none \\ \text{ArmSwivel} = out \\ \text{ForearmHandOrientation} = palm\ towards\ self \\ \text{ArmTouch} = trunk \end{bmatrix} \\ \text{SHOULDER} & = \text{raise shoulders} \\ \text{TRUNK} & = \text{lean froward} \end{bmatrix}$	$\begin{bmatrix} \text{LEFT ARM} & \begin{bmatrix} \text{ArmHeight} = waist \\ \text{ArmDistance} = touch \\ \text{ArmRadialOrientation} = behind \\ \text{ArmRadial Z} = none \\ \text{ArmSwivel} = out \\ \text{ForearmHandOrientation} = palm\ towards\ self \\ \text{ArmTouch}\ trunk \end{bmatrix} \end{bmatrix}$
	$RIGHT\ LEG = \begin{Bmatrix} legrole = normal \\ legshape = closed \\ legposition = forward \\ kneeshape = normal \\ footorientation = opened \\ footshape = flat\ on\ floor \end{Bmatrix}$	$LEFT\ LEG = \begin{Bmatrix} legrole = normal \\ legshape = closed \\ legposition = normal \\ kneeshape = normal \\ footorientation = normal \\ footshape = flat\ on\ floor \end{Bmatrix}$
Attending 	$\begin{bmatrix} \text{RIGHT ARM} & \begin{bmatrix} \text{ArmHeight} = thigh \\ \text{ArmDistance} = \text{touch} \\ \text{ArmRadialOrientation} = \text{none} \\ \text{ArmRadial Z} = downward \\ \text{ArmSwivel} = out \\ \text{ForearmHandOrientation} = \text{palm towards self} \\ \text{ArmTouch} = nottouching \end{bmatrix} \\ \text{SHOULDER} & = raise\ shoulders \\ \text{TRUNK} & = none \end{bmatrix}$	$\begin{bmatrix} \text{LEFT ARM} & \begin{bmatrix} \text{ArmHeight} = chest \\ \text{ArmDistance} = touch \\ \text{ArmRadialOrientation} = front \\ \text{ArmRadial Z} = none \\ \text{ArmSwivel} = out \\ \text{ForearmHandOrientation} = palm\ towards\ self \\ \text{ArmTouch} = trunk \end{bmatrix} \end{bmatrix}$
	$RIGHT\ LEG = \begin{Bmatrix} legrole = normal \\ legshape = apart \\ legposition = backward \\ kneeshape = normal \\ footorientation = opened \\ footshape = flat\ on\ floor \end{Bmatrix}$	$LEFT\ LEG = \begin{Bmatrix} legrole = normal \\ legshape = normal, \\ legposition = normal \\ kneeshape = normal \\ footorientation = normal \\ footshape = flat\ on\ floor \end{Bmatrix}$
Disappear from view 	$\begin{bmatrix} \text{RIGHT ARM} & \begin{bmatrix} \text{ArmHeight} = thigh \\ \text{ArmDistance} = \text{touch} \\ \text{ArmRadialOrientation} = \text{out} \\ \text{ArmRadial Z} = \text{downward} \\ \text{ArmSwivel} = \text{touch} \\ \text{ForearmHandOrientation} = \text{palm towards self} \\ \text{ArmTouch} = \text{nottouching} \end{bmatrix} \\ \text{SHOULDER} & = \text{lower shoulders} \\ \text{TRUNK} & = \text{lower trunk} \end{bmatrix}$	$\begin{bmatrix} \text{LEFT ARM} & \begin{bmatrix} \text{ArmHeight} = \text{abdomen} \\ \text{ArmDistance} = \text{touch} \\ \text{ArmRadialOrientation} = \text{inside} \\ \text{ArmRadial Z} = none \\ \text{ArmSwivel} = \text{touch} \\ \text{ForearmHandOrientation} = \text{palm towards self} \\ \text{ArmTouch} = \text{arm} \end{bmatrix} \end{bmatrix}$
	$RIGHT\ LEG = \begin{Bmatrix} legrole = light \\ legshape = normal \\ legposition = backward \\ kneeshape = normal \\ footorientation = opened \\ footshape = toes \end{Bmatrix}$	$LEFT\ LEG = \begin{Bmatrix} legrole = support \\ legshape = normal \\ legposition = normal, forward, backward \\ kneeshape = slightly\ flexe \\ footorientation = normal \\ footshape = flat\ on\ floor \end{Bmatrix}$

Exuberant



$$\begin{bmatrix} \text{RIGHT ARM} & \begin{bmatrix} \text{ArmHeight} = \text{chest} \\ \text{ArmDistance} = \text{far} \\ \text{ArmRadialOrientation} = \text{side} \\ \text{ArmRadial Z} = none \\ \text{ArmSwivel} = \text{out} \\ \text{ForearmHandOrientation} = palm\ inwards \\ \text{ArmTouch} = \text{notouching} \end{bmatrix} \\ \text{SHOULDER} & = \text{raise shoulders} \\ \text{TRUNK} & = \text{raise trunk} \end{bmatrix}$$

$$\begin{bmatrix} \text{LEFT ARM} & \begin{bmatrix} \text{ArmHeight} = \text{chest} \\ \text{ArmDistance} = \text{far} \\ \text{ArmRadialOrientation} = \text{side} \\ \text{ArmRadial Z} = none \\ \text{ArmSwivel} = \text{out} \\ \text{ForearmHandOrientation} = palm\ inwards \\ \text{ArmTouch} = \text{nottouching} \end{bmatrix} \end{bmatrix}$$

$$RIGHT\ LEG = \begin{Bmatrix} legrole = normal \\ legshape = apart \\ legposition = backward \\ kneeshape = normal \\ footorientation = opened \\ footshape = flat\ on\ floor \end{Bmatrix}$$

$$LEFT\ LEG = \begin{Bmatrix} legrole = normal \\ legshape = normal \\ legposition = normal \\ kneeshape = normal \\ footorientation = normal \\ footshape = flat\ on\ floor \end{Bmatrix}$$

III.3 **Thesis Contribution: Evaluating the static postural expressions of action tendencies**

A previous experiment using the PERMUTATION corpus showed that people reliably assign some action tendencies to the video clips (Clavel and Martin 2009).

In the study that we describe below, we explore if people are able to perceive the action tendencies conveyed by postures that were annotated in the PERMUTATION corpus and replicated in pictures of a virtual character.

III.3.1 Participants

20 subjects (7 female, 13 male, aged 21-60, 79% European, 16% African, and 5% Asian) completed a paper questionnaire and supplied information about age, gender and culture.

III.3.2 Procedure

A questionnaire was set up to assess the extent to which subjects recognize the action tendencies in our pictures of postures. The questionnaire contains two parts.

In the first part of the questionnaire we asked subjects to select the pictures of postures that match the written description of each action tendency. The written descriptions are those proposed by Frijda. The order of presentation of the action tendencies was randomized. For each action tendency, we showed 6 images: 2 target images (i.e. supposed to characterize the action tendency specified in the written description), and 4 distracting images (i.e. neutral or Bodily expressions of other action tendencies).

Hypothesis. We expected that subjects would assign the correct action tendency to each posture, since the design of the posture was informed by the videos of the PERMUTATION corpus that were well recognized in terms of action tendencies.

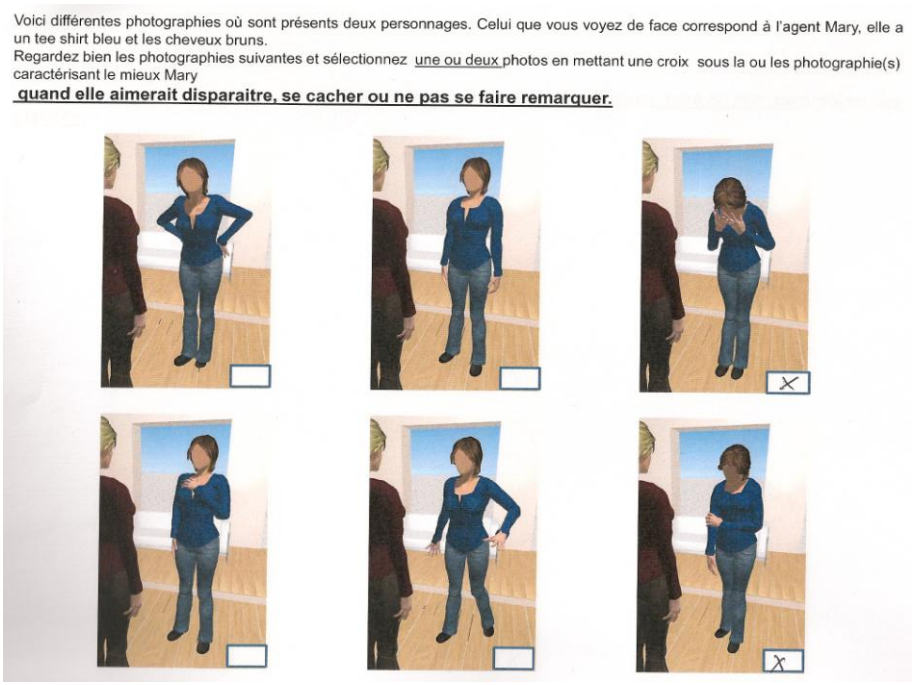


Figure 33: An example of the questionnaire for the recognition of action tendencies

In the second part of the questionnaire, subjects were asked to assign one emotional label to each of the 12 postures. For each posture, they had to select one label out of the following list: Sadness, Joy, Anger, Anxiety, Surprise, Fear, Irritation, Shame, Contempt, Guilt, Disgust, Pleasure, Despair, Pride. We selected these emotion categories because they are the categories that were used by Frijda in his study about the relation between action tendencies and emotion categories (Clavel and Martin 2009).

This forced-choice method was used in several studies about emotional perception and it was observed that subjects agree above chance levels (Coulson 2004), (De Silva and Bianchi-Berthouze 2004), (Ekman 1965), (Ekman and Friesen 1967), (Pitterman and Nowicki Jr 2004),(Wallbott 1998).

Hypothesis. Since different action tendencies correspond to different emotions (Frijda, Kuipers et al. 1989), we expected that our results would be consistent with predictions drawn from psychology studies about action tendencies: subjects should assign emotion categories that are consistent with the action tendencies.

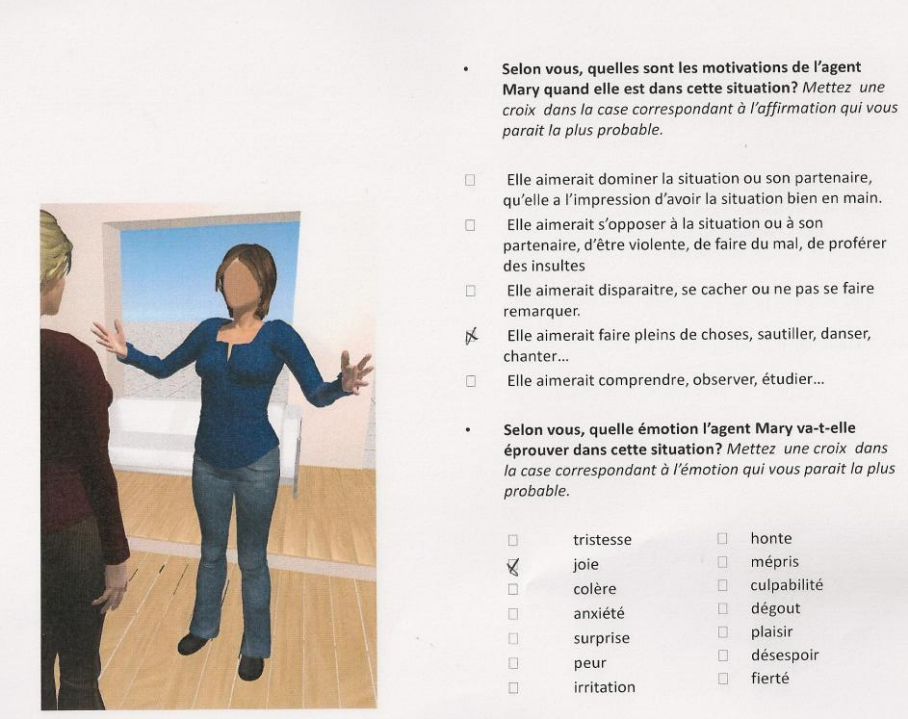


Figure 34: An example of the questionnaire for the attribution of emotion categories

III.3.3 Results

We analyzed 1) whether the subjects recognized the two target postures that we designed for each action tendencies, and 2) how they attributed emotion categories to the 12 postures.

Action Tendency Perception. Three action tendencies (*Attending*, *Disappear from view*, *Exuberant*) had success rates above 50% for one of their two images (the threshold for selecting one image among six would have been 16,7%). For example, among the postures perceived by the subjects as expressing the action tendency Attending, posture #1 was selected 52% of the times. The postural expressions designed for the action tendency In command was not correctly recognized (69% of the answers referred to other postures that the two target images). Table 47 provides for each action tendency the attribution rate for the different postures.

Table 47. Attribution rates for target and distracting postures for each action tendency.

	Posture #1	Posture #2	Distracters (other postures than targets)	TOTAL
Attending	**52%**	10%	38%	100%
Antagonistic	**39%**	27%	33%	100%
Disappear from view	**67%**	33%	0%	100%
Exuberant	**54%**	21%	25%	100%
In command	25%	6%	69%	100%

Emotion Attribution. Table 48 provides the attribution rates of emotion categories for each posture. For each of the postural expressions, we only considered the two highest rates that are above the chance (8.3%). Our results about the assignation of emotion categories to postural expressions of action tendencies are consistent for at least one postural expressions with the results reported by Frijda. In addition, the most of postural expressions that were best recognized in terms of action tendency received emotional attributions that are more consistent with the prediction by Frijda (Frijda, Kuipers et al. 1989).

The table below lists the attribution rates of emotion categories with respect to the two postural expressions of the five action tendencies. The rates in grey shading refer to that is predicted by Frijda. The rates in bold refer to the highest rates per column and that are above the chance (8.3%). This table allows comparing our results to the predictions by Frijda.

Disappear from a view

Regarding the action tendency Disappear from a view, subjects mainly attributed shame and guilt. These two emotions are very close to each other in the Circumplex model of Russell (Russell 1980). Meanwhile, shame refers to the intention to "disappear from view" in Frijda's study. In our study, the subjects also perceived the postures designed for this action tendency as conveying sadness, shame and guilt.

Exuberant

For the action tendency Exuberant, the emotions joy and pride were selected by the subjects. These are all positive emotions. We observed the same results as predicted by Frijda: positive emotion categories (pride, happy) were associated to the action tendency Exuberant.

Attending

Similar results were observed for Attending. The subjects ascribed negative emotion categories (anger, irritation) and surprise. These results showed that subjects assign related emotions to postures that are supposed to express the action tendency Attending by Frijda.

Antagonist

For the postural expressions designed for the action tendency Antagonistic, we found the same pattern of results. Subjects mainly ascribed irritation, contempt, anger, and anxiety as predicted by Frijda.

In command

For the action tendency In command, we observed different results for the two target animations compared to what had been reported in the previous section (recognition of action tendencies). The first target animation was perceived as expressing Pleasure and Pride, which is contradictory to the predictions of Frijda. The second target image was perceived as expressing Anger and Irritation, which is consistent to the predictions by Frijda. The second target posture designed for this action tendency was then more reliably judged in terms of emotion categories than the first target posture. This result is in contrary with the findings in the part of assigning action tendencies, which showed that the first target posture was better perceived in terms of the action tendency In command.

Table 48. Attribution of emotion categories to the postural expressions of action tendencies. The emotions with a level of attribution of more than 10% for each posture are in bold. The percentages for the emotion categories that are predicted by Frijda 1989 for each action tendency are displayed in grey shading.

	Action Tendencies									
	ATTENDING		ANTAGONISTIC		DISAPPEAR		*EXUBERANT*		IN COMMAND	
Emotion Categories	Target Posture #1	Target Posture #2	Target Posture #1	Target Posture #2	Target Posture #1	Target Posture #2	Target Posture #1	Target Posture #2	Target Posture #1	Target Posture #2
sadness	**10%**	0%	0%	0%	**18%**	30%	5%	0%	5%	0%
joy	0%	5%	0%	0%	0%	0%	**47%**	5%	0%	0%
anger	0%	**62%**	**30%**	**15%**	0%	4%	5%	**9%**	0%	**41%**
anxiety	**19%**	0%	0%	25%	0%	0%	0%	**9%**	**20%**	9%
surprise	**33%**	5%	**10%**	**15%**	0%	4%	5%	5%	5%	14%
fear	5%	5%	0%	5%	0%	4%	0%	0%	5%	0%
irritation	5%	**24%**	**25%**	**15%**	0%	**13%**	0%	**9%**	**10%**	**14%**
shame	5%	0%	0%	5%	**55%**	**9%**	0%	0%	10%	0%
contempt	0%	0%	**20%**	5%	0%	4%	**11%**	5%	5%	5%
guilt	5%	0%	0%	0%	**18%**	**17%**	0%	0%	5%	0%
disgust	0%	0%	5%	0%	5%	0%	0%	0%	0%	0%
pleasure	5%	0%	0%	5%	0%	0%	0%	**27%**	**15%**	**9%**
despair	5%	0%	0%	0%	5%	9%	0%	0%	5%	0%
pride	**10%**	0%	10%	**10%**	0%	4%	**26%**	**32%**	**15%**	**9%**

III.3.4 Conclusion

The goal of the study described in this section was to validate the postural expressions of action tendencies that we designed with the MARC virtual agent platform and that were inspired from a video corpus.

These results validate some of the postural expressions of action tendencies that we designed. We observe two main results:

1) One posture for each of three action tendencies received a high recognition rate (Attending, Disappear from view, Exuberant). These postures and their EXPO description are listed in Table 45.

2) The action tendency for which the error rate is the highest (In command) corresponds to postures for which subjects attributed few but different emotions categories.

The results also suggest that the two parts of the questionnaire provide complementary information and are relevant from a methodological point of view. Future studies should be conducted to better understand the similarity of the postures designed for each action tendency and to possibly integrate the best relevant features of the two postures in a single posture representing the action tendency.

The fact that the recognition of an action tendency is not good might explain why the emotional attributions are not consistent with those observed by Frijda. Although one posture was well recognized for several action tendencies (*Attending*, *Disappear from view*, *Exuberant*), the other postures designed for these action tendencies were not well recognized. This might be explained by the fact that we asked subjects to select one or two images representing the action tendency, which means that the second response was optional.

In addition, we observed that the number of PERMUTATION annotated clips used to inform the design of the postural expressions had an effect on the recognition performance. The three action tendencies that received good recognition performance (*Attending*, *Disappear from view*, *Exuberant*) were based on a higher number of clips than the action tendencies that were not well recognized (In command and Antagonistic). A high number of annotated clips might thus provide more information and improve the accuracy of the designed postures.

The postures that were best recognized for each action tendency were kept for the study described in the next section which explores how action tendencies can be conveyed by dynamic animations.

III.4 Thesis Contribution: Designing dynamic postural expressions of action tendencies

In the previous section we explained how we designed and evaluated static postural expressions (e.g. pictures) of action tendencies. The related work section pointed the importance of the dynamics in the expression and perception of emotions. Few studies explored the dynamics of postural expressions of the action tendency component of emotions.

Thus, designing animations intended to express action tendencies raises several questions: How to specify the dynamics of the animations of the static pictures expressing action tendencies that we presented in the previous section? How long should last an animation: how long and how explicit should be the postural expression of an action tendency? Are there any differences between the different action tendencies in this respect?

III.4.1 Relations between action tendencies and movement quality

Several authors have proposed various definitions of the different features of movement quality that might be related to the expression of emotion (Mancini and Pelachaud 2007). In a previous study, 299 subjects reported different features of movement quality that they perceived in the 100 videos of the PERMUTATION corpus (Clavel and Martin 2009). These features of movement quality included the intensity of the movement and its temporal extent. Intensity is related to the amount of energy and tension. Temporal extent is related to the duration of the movement and is thus inversely related to its speed.

We drew an ANOVA on the results of the PERMUTATION study to infer the relations between these two parameters of movement quality and the action tendencies.

The only significant result that we found was that the videos perceived as expressing the *Antagonistic* action tendency significantly had higher intensity ($p=0.008$) and lower temporal extent ($p=0.01$) than the videos featuring the *Disappear from view* action tendency. In order to have values of intensity and temporal extent for each action tendency, we also use descriptive results and the original videos.

The resulting mapping is provided in Table 49. The next section explains how we used this mapping for designing dynamic animations of the static postures that were validated in the previous section.

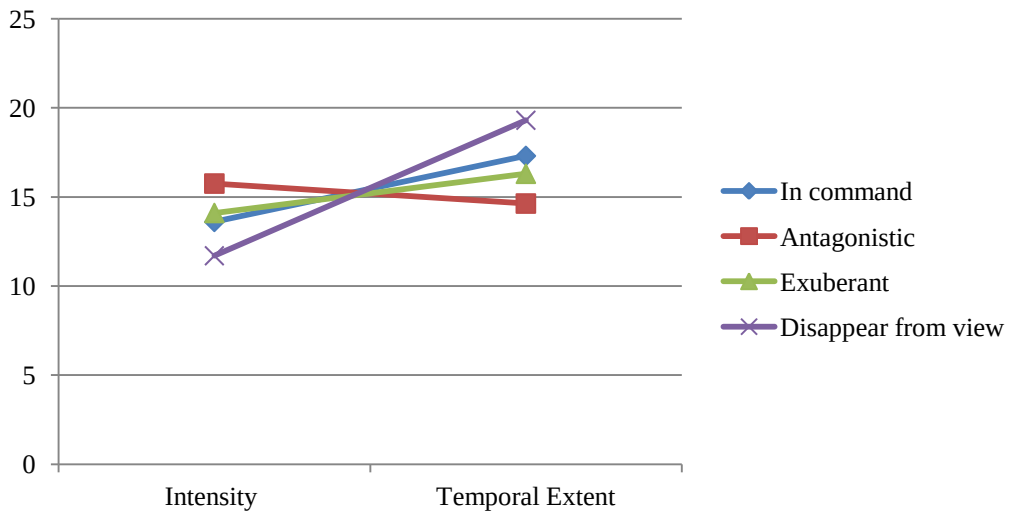


Figure 35: Relations between the four action tendencies (IN: in command, AG: Antagonistic**, EX:** Exuberant**, DFV:** Disappear from View**) on the scales of intensity of movement and temporal extent**

Table 49: Mapping between action tendencies, intensity of movement and temporal extent

Action tendencies	*Intensity of movement*	*Temporal extent*
Disappear from view	low	high
In command	medium	medium
Antagonistic	high	low
Exuberant	medium	medium
Attending	medium	low

III.4.2 Specifying the dynamics of MARC animations of action tendencies

We kept the static postural expressions of action tendencies that were best recognized in the study described in the previous study. In this section, we explain how we designed their dynamic animation using the MARC platform.

We inspired from the results on the PERMUTATION corpus (Figure 35) to inform the values of Intensity of movement and Temporal extent for the animations of our virtual character.

In addition, we also used the notion of *opposing force* that animators use to make a movement appear realistic (Stanchfield and Hahn 2009). In Physics, the opposing force refers to an agent that, if unfettered by equally opposing forces, will cause a net acceleration in the object upon which it is acting. An opposite force will thus cause an acceleration of the movement. People usually involve an opposing force before every movement to accelerate the movement. We kept this notion of opposing force when converting our pictures of postural expressions of action tendencies into animations. We inserted an opposing force frame after the starting default keyframe and before the keyframe that expresses the action tendency (called hereafter the stroke keyframe).

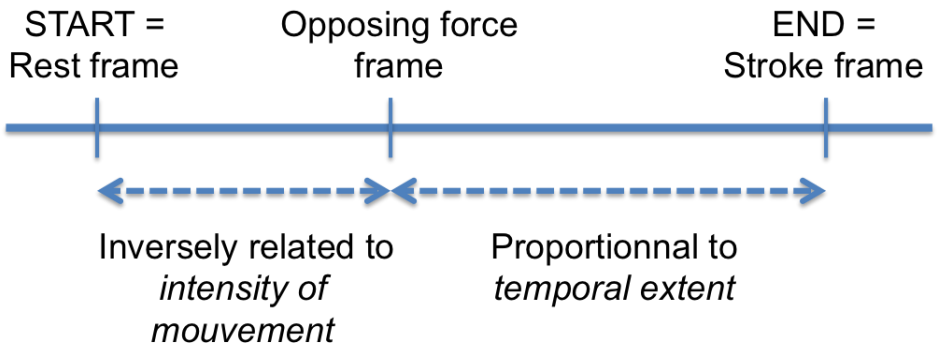


Figure 36: Combining opposing force, intensity of movement and temporal extent for specifying the dynamics of our postural expressions of action tendencies

An animation starts by a resting keyframe, followed by an opposing force frame, and ends with a stroke frame (Figure 36). We decided to stop our animations with the stroke frame and we did not include any retraction phrase because we consider our expressions as a preparation to action rather than a full animation of the action itself.

Once the opposing force keyframes were created, we parameterized the MARC animations by adjusting the timing according to the different action tendencies. We expressed a high intensity by a short inbetween duration between the resting frame and the opposing force

frame. A high level of temporal extent led to a long duration between the opposing force frame and the final stroke frame.

The values for all the timing parameters are summarized in Table 50. We applied a default duration for the whole animation of 1000 ms. This duration is similar to the durations of the whole videos (960 ms) used in a similar study about the perception of dynamics of facial expressions of emotions (Barkhuysen, Krahmer et al. 2010). We then selected values for 1) the duration of the part of the animation before the opposing force frame, and 2) the duration of the animation between the opposing force frame and the final stroke frame.

Table 50: Timing parameters for the animations of the action tendencies

Action tendency	***Resting frame (starting)***	*In-betweens (ms)*	*Opposing force frame*	*In-betweens (ms)*	*Stroke frame (ending)*
Attending		160		1080	
In command		300		900	
Exuberant		280		940	
Antagonistic		100		1500	
Disappear from view		2360			
Neutral		1000			

III.5 **Thesis contribution: Evaluating the dynamic postural expressions of action tendencies**

We conducted a 2nd experimental study to evaluate how our dynamic bodily expressions of action tendencies were perceived.

Our first goal was to validate the animation parameters that we described in the previous section.

Hypothesis. Our first assumption was that the animations using our mapping between intensity of movement, temporal extent and action tendencies would lead to a better recognition of the action tendencies than the animations that do not use these parameters.

We were also willing to investigate the impact of the duration of the animation on the recognition of the action tendency. We used a gating design similar to (Barkhuysen, Krahmer et al. 2010) in which clips stimulus are presented in segments of increasing duration and subjects are asked to propose the word being presented and to give a confidence rating after each segment. We aimed to investigate what is the recognition speed for dynamic bodily expressions to action tendencies and categorical emotions? In other words, is there an absolute threshold for identifying states of readiness to execute a given kind of action from the humans?

Hypothesis. Our second assumption was that during social interaction the perception of the action tendency would need short time during which the action tendency has not yet been totally performed.

In summary, the goals of the experiment described in this section were:

- Validating the mapping model of dynamic bodily expressions of action tendencies;
- Investigating the timing that would convey the best the studied action tendencies.

The results of this experiment were expected to be useful to the design of expressive virtual characters.

III.5.1 Stimuli

The design process followed the next steps.

First, we specified 2 versions of dynamic animations for each of the 5 action tendencies. The first version is parameterized (section III.2). The second version is non-parameterized: use the neutral parameters (section III.2).

Second, we generated 4 sets of video clips (.mov) at frame rate 200.000 fps for each of the 5 action tendencies. They were cut from 488×720 pixels.

These animations are:

- N3/3: A “neutral” animation using linear interpolation, a full duration of 1000 ms and without any opposing force frame
- Condition 1: P3/3: A full animation (3/3) using our mapping values for movement quality
- Condition 2: P1/3: An animation that contains only the first third of the full animation P3/3
- Condition 3: P2/3: An animation that contains only the two first thirds of the full animation P3/3

Different actions tendencies might lead to different durations of bodily expressions. In our study, the animations based on the video corpus have different durations for different action tendencies. The different gating version (1/3 and 2/3) of the same animation have the same starting resting frame, but have different ending frames. Ending frames correspond to different gate settings.

A total of 20 video clips were generated and used in the study. A female virtual character expressed bodily expressions in the same social scene setting as in the static posture design.

An online questionnaire was set up. Age, gender, cultural origins were required to report before the questionnaire. The aim of the questionnaire was to assess the extent to which subjects recognize the action tendencies in dynamic bodily expressions according to different timing settings. The questionnaire composed of 60 questions. For each of the 20 video clips, we asked 3 questions. The order of presentation of video clips was randomized.

First, we asked subjects to judge their agreement regarding the relevance between a set of verbal descriptions of action tendencies and the related dynamic bodily expressions. The written descriptions are those proposed by Frijda. We used a 5-point Likert scale (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree) to measure the level of agreement. Then, the confidence level was required to report with respect to their judgments throughout a 5-point Likert scale (Extremely confident, Very Confident, Somewhat confident, Not very confident, Not at all confident). Finally, subjects were asked to assign one emotional label to each of the dynamic animations. The list of emotions was as the same as that in the experiment 1 (sadness, joy, anger, anxiety, surprise, fear, irritation, shame, contempt, guilt, disgust, pleasure, despair, pride). Subjects could select multiple answers.

III.5.2 Design

This experiment used a repeated-measures design with Movement Quality as within-subjects factor (with levels: PARAMETERIZED AND NEUTRAL), SEGMENTATION AND DURATION OF VIDEO (with levels: P3/3: condition 1; P1/3 TIME: condition 2; P2/3 TIME: condition 3), and PERCEIVED ACTION TENDENCIES (with levels: ATTENDING, DISAPPEAR FROM VIEW, ANTAGONISTIC, EXUBERANT, IN COMMAND) as the dependent factors.

Table 51 : Factorial design of the 2nd experimental study

		dependant factor 1	dependant factor 2	dependant factor 3
independent factor 1	Movement Quality	PERCEIVED ACTION TENDENCIES (section III.5.3.2.1)	CONFIDENCE (section III.5.3.3)	EMOTION ATTRIBUTION (section III.5.3.4)
independent factor 2	SEGMENTATION AND DURATION OF VIDEO			

III.5.3 Participants

32 subjects participated in the experiment, 16 female and 16 male, with an average age of 31 (range 22-60), 75% European, 6.25% African, 9.38% Asian and 9.38% American. Each subject viewed 20 animations and evaluated them in terms of action tendency and emotion attributions. This kind of design requires fewer participants, and allowed us to monitor the effect upon individual easily and lower the possibility of individual differences skewing the results. However, the within subject design over between subject design lowers the chances of subjects suffering boredom after viewing a long series of subtly differentiated animations.

III.5.3.1 Brief summary of main results

This section briefly summarizes the main results. All the results are described in details in the next sections.

Impact of the quality of movement. Table 52 summarizes the results regarding the quality of movement. Our parameters of movement quality had neither impact on the reported action tendency nor on the confidence of subjects in their ratings of action tendencies. Yet, we observed that our parameters of movement quality had an impact on the attribution of emotion categories to our animations. The animations using our mapping for the intensity of movement and temporal extent received emotion categories that are closer to the emotion categories proposed by Frijda.

Table 52: Validation of the effects of independent factor Movement Quality on the different dependant factors

Independent factor 1	Dependant factor 1	Dependant factor 2	Dependant factor 3
	PERCEIVED ACTION TENDENCIES	CONFIDENCE	EMOTION ATTRIBUTION
Movement Quality	No effect observed	No effect observed	Animations with movement quality more consistent with Frijda

Impact of segmentation and duration of the animation. Table 53 summarizes the results regarding the impact of the segmentation and duration of the animations on reported action tendencies, confidence and emotion categories. The segmentation and duration of the animations had an impact on the perceived action tendencies and confidence: the recognition rate of action tendencies and the confidence of subjects were lower in the 1/3 condition than in the 2/3 and 3/3 conditions.

The segmentation and duration of the animations also had an impact on the attribution of emotion categories to the animations: the attribution of emotion categories was more consistent with the observations reported by Frijda in the 2/3 condition than in the 1/3 and 3/3 conditions.

Table 53: Validation of the effects of independent factor Segmentation and Duration of the Animation on the different dependant factors

Independent factor 2	**Dependant factor 1**	**Dependant factor 2**	**Dependant factor 3**
	PERCEIVED ACTION TENDENCIES	CONFIDENCE	EMOTION ATTRIBUTION
SEGMENTATION AND DURATION OF THE ANIMATION	Lower in 1/3 than 2/3 and 3/3	Lower in 1/3 than 2/3 and 3/3	More consistent with Frijda in 2/3 than in 1/3 and 3/3

The following sections detail these results.

III.5.3.2 Results on Perceived Action Tendencies

Evaluations of perceived action tendencies were subjected to repeated measures ANOVA. In ANOVA tests, the F-ratio is used to compare the variance between the groups to the variance within the groups. The larger the F-ratio, the more certain we are that there is a difference between the groups. We have a critical alpha level at the .05. If the probability of the F-ratio is less than or equal to .05, it indicates that a significant difference exists between at least two of groups. Nonetheless, the F-ratio does not tell which groups are different from the others. We then need to run multiple comparison tests (post hoc analyses) to enable to investigate which pairs of conditions / action tendencies are significantly different as well as the interaction effects between experimental conditions and action tendencies. Post hoc analyses were performed with the Scheffé method. The figures of multiple comparison tests showed the estimates with comparison intervals around them. Groups are significantly different if their intervals are disjoint. They are not significantly different if their intervals overlap.

Based on these statistical methods and procedures, we aimed to investigate:

- whether the means of evaluations scores of ACTION TENDENCIES are significantly different across levels of Movement Quality;
- whether the means of evaluations scores of ACTION TENDENCIES are significantly different across levels of Segmentation and Duration of the Animation
- which pairs of levels of Segmentation and Duration of the Animation are significantly different;
- which pairs of levels of ACTION TENDENCIES are significantly different across levels of Movement Quality ;
- which pairs of levels of ACTION TENDENCIES are significantly different across levels of Segmentation and Duration of the Animation ;
- whether there is an interaction between ACTION TENDENCIES and Movement Quality;
- whether there is an interaction between ACTION TENDENCIES and Segmentation and Duration of the Animation.

We coded the evaluation scores of action tendencies as a value of 0 (Strongly disagree), 1 (Disagree), 2 (Neither agree nor disagree), 3 (Agree), 4 (Strongly agree). We presented the results according to the effect of independent factors.

III.5.3.2.1 Effect of Movement Quality on Perceived Action Tendencies

We performed a one-way ANOVA, with Movement Quality as within-subjects factors and perceived action tendencies as the dependent variable. The main effect of Movement Quality had no significance on perceived action tendencies, $F(1,310) = 0.698$, $MS = 0.2$, $p > .05$.

Evaluation scores of action tendencies were significantly different each other across overall levels of Movement Quality, $F(4, 310) = 27.05$, $p<.05$. Post hoc analyses showed that the mean score of Disappear from View ($M= 3.1406$, $SD= 0.1439$) was significantly higher than

that of Attending (M= 2.3906, SD= 0.1439), Antagonistic (M= 2.5625, SD= 0.1439), and Exuberant. (M= 1.1719, SD= 0.1439). The mean score of Attending (M= 2.3906, SD= 0.1439), Antagonistic (M= 2.5625, SD= 0.1439) and In command (M= 2.7969, SD= 0.1439) was significantly higher than that of Exuberant (M= 1.1719, SD= 0.1439).

III.5.3.2.2 Effect of Segmentation and Duration of the Animation on Perceived Action Tendencies

The main effect of Segmentation and Duration of the Animation was significant on perceived action tendencies, F (2,465) = 16.15, MS = 21.306, p < .05. Post hoc analyses showed that the mean score of overall action tendencies was significantly lower for condition 3 (M=1.7188, SD=0.091) than for condition 1 (M=2.3875, SD=0.091) and condition 4 (M=2.3062, SD=0.091). The results of pairwise comparison of levels of Segmentation and Duration of the Animation were presented in the table below. It compared 3 levels of Segmentation and Duration of the Animation two by two in the form of a 3-column matrix. It indicated the estimated difference in means and a confidence interval of the difference. For example, the mean of condition 1 minus the mean of condition 2 is estimated to be-0.6688, and a 95% confidence interval for the true difference of the means is (0.3678, 0.9697).

Evaluation scores of action tendencies were significantly different each other across overall levels of Segmentation and Duration of the Animation, F (4,465) =22.57, MS=29.778, p<.05. We then applied a multiple comparison test (post hoc analysis) to reveal which pairs of action tendencies were significantly different. The mean score of Attending (M= 2.375, SD=0.117), Disappear from View (M=2.740, SD=0.117) and In command (M=2.438, SD=0.117) was respectively significantly higher than that of Antagonistic (M=1.771, SD=0.117) and Exuberant (M=1.365, SD=0.117). The results of Post hoc analyses were sum up in Table 54. Table 54 listed the results of the test comparing 5 action tendencies two by two in the form of a five-column matrix. It indicated the estimated difference in means and a confidence interval of the difference. For example, the mean of Attending minus the mean of Disappear from View is estimated to be -0.3646, and a 95% confidence interval for the true difference of the means is (-0.8168, 0.0877).

Table 54: Pairwise comparison of 5 action tendencies: the estimated difference in means and a confidence interval of the difference

	AD	DFV			AG			EX			IN		
AD		-0.81	-0.36	0.09	0.16	0.60	1.06	0.56	1.01	1.46	-0.52	-0.06	0.39
DFV					0.52	0.97	1.42	0.92	1.37	1.83	-0.15	0.30	0.75
AG								-0.05	0.40	0.85	-1.12	-0.67	-0.21
EX											-1.53	-1.07	-0.62
IN													

We performed univariate analyses within each level of Segmentation and Duration of the Animation, with evaluation scores of action tendencies as the dependent variable.

Within the condition 1, evaluation scores of action tendencies were significantly different each other, $F(4, 155) = 14.88$, $MS = 18.306$, $p<.05$. The mean score of Disappear from View (M= 3.2188, SB=0.1961) was significantly higher than Attending (M= 2.4063, SB= 0.1961), Antagonistic (M= 2.3438, SB=0.1961) and Exuberant (M= 1.1875, SB=0.1961). The mean score of Attending (M= 2.4063, SB= 0.1961), Disappear from View (M= 3.2188, SB=0.1961), Antagonistic (M= 2.3438, SB=0.1961) and In command (M= 2.7813, SB=0.1961) was respectively significantly higher than that of Exuberant (M= 1.1875, SB=0.1961).

Within the condition 2, $F(4, 155) = 5.41$, $MS= 7.234$, $p<.05$. Post hoc analyses showed that the mean score of perceived Attending (M=2.3438, SD=0.2045) was significantly higher than Antagonistic (M=1.093, SD=0.2045) and Exuberant (M=1.438, SD=0.205). The mean score of In command (M=1.9063, SD=0.205) was significantly higher than that of Antagonistic (M=1.093, SD=0.2045).

Within the condition 3, $F(4, 155) =16.23$, $MS = 14.163$, $p<.05$. The mean score of Disappear from View (M= 3.188, SD=0.208) was significantly higher than that of Attending (M= 2.375, SD= 0.208), Antagonistic (M=1.8750, SD=0.208) and Exuberant (M=1.4688, SD=0.208). The mean score of In command (M=2.6250, SD=0.208) was significantly lower than that of Exuberant (M=1.4688, SD=0.208).

The interaction of Segmentation and Duration of the Animation and perceived action tendencies was significant, $F(8, 465) = 3.76$, $MS = 4.963$, $p < .05$, indicating that the effect of Segmentation and Duration of the Animation was different from some action tendencies to others. We performed univariate analyses within each of action tendencies, with Segmentation and Duration of the Animation as within-subjects factor and evaluation scores of action tendencies as the dependent variable.

For the action tendency Attending, no significant difference was found among levels of Segmentation and Duration of the Animation, $F(2, 93) =0.03$, $MS=0.03$, $p>.05$. The scores of perceived action tendencies were relatively high across levels of Segmentation and Duration of the Animation for both condition 1 (M=2.406, SB =0.176), condition 2 (M=2.344 SB =0.176) and condition 3 (M=2.375, SB =0.176).

For the action tendency Disappear from View, $F(2, 93) =13.56$, $MS= 20.64$, $p<.05$, the mean scores for condition 2 (M=1.813, SB = 0.218) was significantly lower than for condition 1 (M=3.219, SB =0.218) and condition 3 (M=3.188, SB =0.218). No significant difference was found between condition 1 and condition 3.

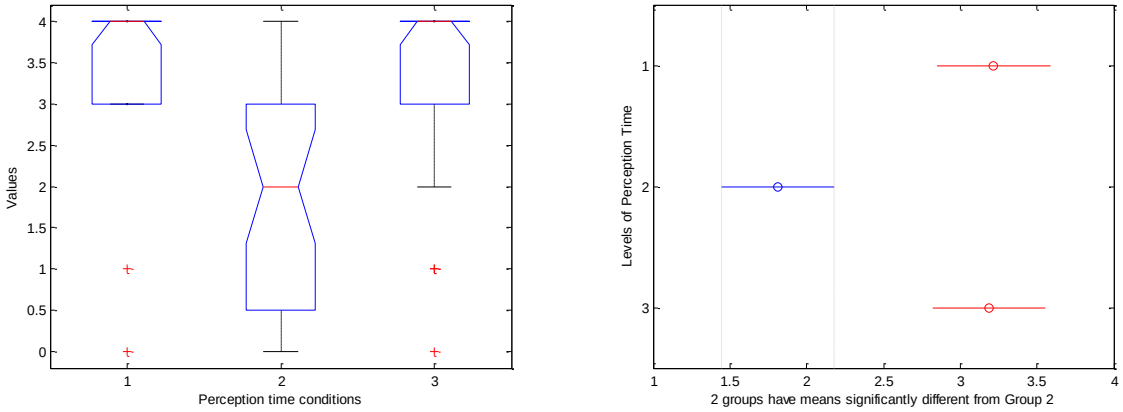


Figure 37: Effect of Segmentation and Duration of the Animation within the action tendency Disappear from View **(comparison of mean scores of perceived action tendencies).**

For the action tendency Antagonistic, F (2, 93) =10.65, MS= 12.76, p<.05, the mean scores for condition 2 (M=1.094, SB = 0.194) was significantly lower than for condition 1 (M=2.344, SB =0.194) and condition 3 (M=1.875, SB =0.194).

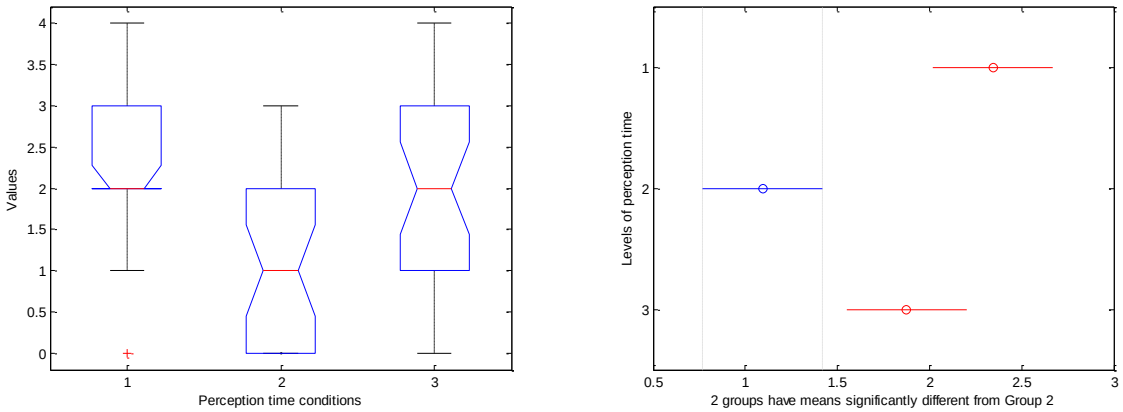


Figure 38: Effect of Segmentation and Duration of the Animation within the action tendency Antagonistic **(comparison of mean scores of perceived action tendencies).**

For the action tendency Exuberant, no significant difference was found among levels of Segmentation and Duration of the Animation, F (2, 93) =0.52, MS=0.760, p>.05. The scores of perceived action tendencies were generally low across levels of Segmentation and Duration of the Animation for both condition 1 (M=1.188, SB =0.213), condition 2 (M=1.438, SB =0.213) and condition 3 (M=1.469, SB =0.213).

For the action tendency In command, F (2, 93) =4.85, MS= 6.969, p<.05, the mean scores for condition 2 (M=1.906, SB =0.2119) was significantly lower than that for condition 1 (M=2.781, SB =0.2119) and condition 3 (M=2.625, SB =0.2119).

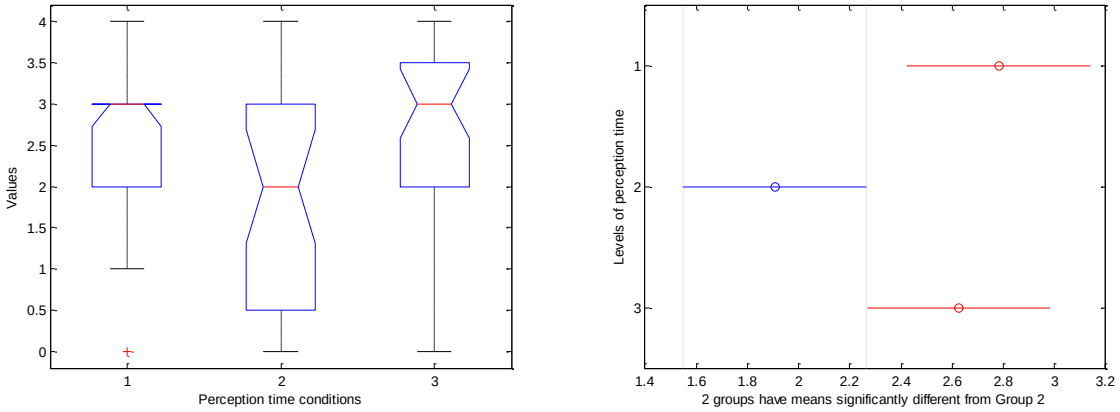


Figure 39: Effect of Segmentation and Duration of the Animation within the action tendency in command (comparison of mean scores of perceived action tendencies).

III.5.3.2.3 Summary

The section III.5.3.2 reported the results investigating the effect of Movement Quality and the effect of Segmentation and Duration of the Animation on the perceived action tendencies.

The effect of Movement Quality was not found on perceived action tendencies. At both levels of Movement Quality (P3/3 and N3/3): 1) Disappear from View was better perceived than Attending, Antagonistic and Exuberant; 2) Attending, Antagonistic and In command were better perceived than Exuberant.

The effect of Segmentation and Duration of the Animation was found on perceived action tendencies: 1) action tendencies were better perceived when presenting a 2/3 time or a 3/3 animation than when presenting a 1/3 time animation to subjects; 2) The action tendency Attending, Disappear from View and In command were better perceived than Antagonistic and Exuberant across overall levels of Segmentation and Duration of the Animation.

Concretely, when presenting a full 3/3 animation to subjects, Disappear from View was better perceived than Attending, Antagonistic and Exuberant. *Attending*, Disappear from View, Antagonistic, In command were better perceived than Exuberant. When presenting 1/3 time animation to subjects, Attending was significantly better perceived than Antagonistic and Exuberant. In command was significantly better perceived than Antagonistic. When presenting 2/3 time animation to subjects, Disappear from View was better perceived than Attending, Antagonistic and Exuberant. In command was better perceived than Exuberant.

In addition, an interaction effect was found with respect to Segmentation and Duration of the Animation and perceived action tendencies. When presenting 2/3 time animation or 3/3 animation to subjects, 3 action tendencies were better perceived than when presenting 1/3 time animation: Disappear from View, Antagonistic and In command. No significant interaction effect was found for Attending and Exuberant.

III.5.3.3 Results on Perceived Confidence of Evaluations

We coded the evaluation scores of confidence as values 4 (Extremely confident), 3 (Very Confident), 2 (Somewhat confident), 1 (Not very confident), 0 (Not at all confident). We presented the results according to the effect of independent factors. We investigated if confidence scores for the same action tendency are significantly different among levels of Movement Quality and Segmentation and Duration of the Animation.

III.5.3.3.1 Effect of Movement Quality on judgment confidence

No significant difference was found across levels of Movement Quality for the mean scores of confidence, $F(1, 310) = 10.14$, $MS = 0.2$, $p > .05$.

III.5.3.3.2 Effect of Segmentation and Duration of the Animation on Confidence

The main effect of Segmentation and Duration of the Animation was significant on confidence of judgments, $F(2, 465) = 21.26$, $MS = 17.665$, $p < .05$. Post hoc analyses showed that the mean score in condition 2 ($M = 3.2563$, $SD = 0.0721$) was significantly lower than in condition 1 ($M = 3.8937$, $SD = 0.0721$) and in condition 3 ($M = 3.7375$, $SD = 0.0721$).

Confidence scores of action tendencies were significantly different each other across overall levels of Segmentation and Duration of the Animation, $F(4, 465) = 4.97$, $MS = 4.128$, $p < .05$. Post hoc analyses showed that the mean score of Attending ($M = 3.3750$, $SD = 0.0930$) was significantly lower than that of Disappear from View ($M = 3.8854$, $SD = 0.0930$) and Antagonistic ($M = 3.7500$, $SD = 0.0930$). The mean score of Exuberant ($M = 3.4688$, $SD = 0.0930$) was significantly lower than that of Disappear from View ($M = 3.8854$, $SD = 0.0930$).

No interaction effect was found between Segmentation and Duration of the Animation and confidence of judgments on action tendencies, $F(8, 465) = 1.76$, $MS = 1.464$, $p > .05$.

III.5.3.3.3 Summary

The section III.5.3.3 reported the results investigating the effect of Movement Quality and the effect of Segmentation and Duration of the Animation on the confidence of judgements on action tendencies.

The effect of Movement Quality was found on the confidence of judgments on action tendencies. The effect of Segmentation and Duration of the Animation was significant on the confidence of judgments on action tendencies. Generally, when presenting the 1/3 time animation to subjects, their confidence was lower than when presenting the 3/3 animation or 2/3 time animation. *Exuberant* received lower confidence than Disappear from View across overall levels of Segmentation and Duration of the Animation.

III.5.3.4 Results on Emotion Attribution

For this analysis, we recorded the responses such that non-answers in the checklist were mapped to a value of 0, and answers in the checklist were mapped to 1. There were 1142 answers on the 20 checklists. Table 55 shows the proportion of answers as functions of action tendencies and independent factors (Movement Quality and Segmentation and Duration of the Animation).

We investigated:

- whether the attribution rates of emotion categories for the same action tendency were consistent across 4 experimental conditions;
- at which level(s) of Movement Quality / Segmentation and Duration of the Animation, the attribution rates of emotion categories fits the most to the predictions by Frija to others.

III.5.3.4.1 Effect of Movement Quality on Emotion Attribution

Except for the action tendency In command and Exuberant, we observed that the emotional attribution fit better to the predictions by Frijda for the parameterization condition than for the neutral condition.

For the action tendency Attending, subjects mainly ascribed anxiety and surprise in the parameterization condition (compared to anxiety only in neutral condition). For the action tendency Antagonistic, subjects mainly ascribed anger, surprise and irritation in parameterization condition but only anger and irritation in neutral condition. For the action tendency Disappear from View, subjects mainly ascribed sadness, shame in parameterization condition and the neutral condition. No difference was found between these conditions. For the action tendency Exuberant, subjects mainly ascribed irritation in both 2 conditions. Subjects mainly ascribed anger in the parameterization condition but not in the neutral condition. This indicated that in the parameterization condition, subjects attributed the predicted emotional categories that relate to the action tendency Exuberant. For the action tendency In command, subjects mainly ascribed anger and irritation in both 2 conditions. Subjects mainly ascribed surprise in neutral condition which did not fit the prediction by Frijda.

III.5.3.4.2 Effect of Segmentation and Duration of the Animation on Emotion Attribution

For the action tendency Attending, subjects mainly ascribed anxiety and surprise in condition 1 but only anxiety in condition 3. For the action tendency Antagonistic, subjects mainly ascribed anger, surprise and irritation in condition 1 but only anger and irritation in condition 3. For the action tendency Disappear from View, subjects mainly ascribed sadness, shame in condition 1, and 3. However, subjects mainly ascribed guilt in condition 2 and in condition 3, but not in condition 1.For the action tendency Exuberant, subjects mainly ascribed irritation

in condition 1, 2 and 4. Subjects mainly ascribed anger in condition 1 and 3. Subjects mainly ascribed surprise in condition 3 and 4 but not in condition 1 or in condition 2. For the action tendency In command, subjects mainly ascribed anger in all conditions. They mainly ascribed irritation in condition 1, 2 and 4. Subjects mainly ascribed surprise in condition 2 and 3.

The attribution of emotional categories was quite scattered in condition 2 for all the action tendencies nonetheless of the action tendency Exuberant.

III.5.3.4.3 Summary

The section III.5.3.3 reported the results investigating the effect of Movement Quality and the effect of Segmentation and Duration of the Animation on the attribution of emotional categories.

Except for the action tendency Exuberant and Disappear from View, the ascribed emotional categories fit better to the predictions by Frijda in the P3/3 condition than in the N3/3 condition. Moreover, when presenting the 2/3 time animation to subjects, the attribution of emotional categories for overall action tendencies was more consistent to the prediction by Frijda than when presenting the 3/3 animation or the 1/3 time animation.

Table 55 : Attribution of emotion categories to the dynamic bodily expressions of action tendencies (the emotions with the highest level of attribution for each expression are in bold; the predictions by Frijda are in grey shading).

	AD				AG				DFV				EX				IN			
	P	N			P	N			P	N			P	N			P	N		
	C1		C2	C3	C1		C2	C3	C1		C2	C3	C1		C2	C3	C1		C2	C3
sadness	0.00%	4.00%	0.00%	12.00%	0.00%	0.00%	2.50%	0.00%	**25.29%**	**17.65%**	16.67%	**19.28%**	1.22%	1.79%	2.70%	3.45%	0.00%	0.00%	2.13%	0.00%
joy	0.00%	4.00%	8.57%	2.00%	1.85%	3.33%	2.50%	5.88%	1.15%	0.00%	0.00%	1.20%	8.54%	8.93%	5.41%	10.35%	1.89%	1.69%	2.13%	0.00%
anger	2.00%	2.00%	2.86%	4.00%	**24.07%**	**35.00%**	12.50%	**23.53%**	0.00%	0.00%	2.08%	1.20%	**24.39%**	**21.43%**	**18.92%**	13.79%	**30.19%**	**25.42%**	**23.40%**	**31.58%**
anxiety	**18.00%**	12.00%	11.43%	14.00%	3.70%	1.67%	10.00%	5.88%	13.79%	15.29%	14.58%	14.46%	4.88%	1.79%	5.41%	5.17%	1.89%	3.39%	8.51%	7.02%
surprise	**24.00%**	**24.00%**	14.29%	**18.00%**	**20.37%**	11.67%	17.50%	15.69%	0.00%	0.00%	6.25%	0.00%	10.98%	10.71%	**21.62%**	**20.69%**	16.98%	**18.64%**	**21.28%**	12.28%
fear	8.00%	6.00%	0.00%	10.00%	0.00%	0.00%	5.00%	0.00%	9.20%	10.59%	8.33%	10.84%	0.00%	0.00%	10.81%	1.72%	0.00%	0.00%	6.38%	0.00%
irritation	6.00%	8.00%	14.29%	6.00%	**20.37%**	**28.33%**	12.50%	**23.53%**	0.00%	1.18%	4.17%	1.20%	**21.95%**	**28.57%**	13.51%	**22.41%**	**26.42%**	**23.73%**	17.02%	**28.07%**
shame	10.00%	4.00%	8.57%	6.00%	0.00%	0.00%	7.50%	3.92%	**18.39%**	**23.53%**	12.50%	**22.89%**	0.00%	0.00%	5.41%	0.00%	0.00%	0.00%	0.00%	0.00%
contempt	4.00%	2.00%	2.86%	6.00%	9.26%	8.33%	0.00%	3.92%	0.00%	0.00%	2.08%	0.00%	9.76%	10.71%	5.41%	1.72%	15.09%	15.25%	8.51%	8.77%
guilt	10.00%	10.00%	8.57%	8.00%	0.00%	0.00%	12.50%	5.88%	17.24%	16.47%	**20.83%**	**18.07%**	0.00%	1.79%	0.00%	1.72%	0.00%	0.00%	0.00%	0.00%
disgust	0.00%	2.00%	5.71%	0.00%	3.70%	3.33%	0.00%	0.00%	2.30%	1.18%	2.08%	1.20%	3.66%	3.57%	5.41%	1.72%	1.89%	5.08%	4.26%	8.77%
pleasure	6.00%	10.00%	8.57%	2.00%	5.56%	0.00%	12.50%	3.92%	0.00%	0.00%	2.08%	0.00%	4.88%	5.36%	2.70%	6.90%	0.00%	3.39%	2.13%	0.00%
despair	2.00%	0.00%	0.00%	0.00%	0.00%	1.67%	2.50%	1.96%	12.64%	14.12%	8.33%	9.64%	6.10%	3.57%	2.70%	1.72%	0.00%	0.00%	0.00%	0.00%
pride	10.00%	12.00%	14.29%	12.00%	11.11%	6.67%	2.50%	5.88%	0.00%	0.00%	0.00%	0.00%	3.66%	1.79%	0.00%	8.62%	5.66%	3.39%	4.26%	3.51%

III.5.4 Discussion

Recognition speed of action tendencies based on the valency of related emotions

The action tendency Attending was better perceived in the P1/3 condition than in the two other conditions, i.e. P2/3 and P3/3). The action tendency Disappear from View was better perceived in the P2/3 time condition and the P3/3 condition than in the P1/3 condition. Meanwhile, we bear in mind that, the action tendency Disappear from View involved the negative emotions such like sadness, shame and guilt, while the action tendency Attending rather involved neutral emotions such like surprise and anxiety. Hence, the two action tendencies can be distinguished on a valency scale. These results are in line with the valency effect reported by (Leppanen and Hietanen 2004; Barkhuysen, Krahmer et al. 2010), which reported that the recognition speed is faster for positive than negative emotions.

High-level confidence when recognizing threatening bodily expressions

The results about confidence of judgments were mostly similar to those about the perceived action tendencies. Except for that Attending received lower confidence than Antagonistic (this action tendency and received higher scores than Antagonistic in terms of perceived action tendencies). It suggests that even if it might be more difficult to perceive Antagonistic than Attending from the target bodily expressions, users preferred to be certain than lacking of concern face to such threatening bodily expressions. This result might be explained by the fact that humans are best able to detect potentially harmful information (i.e. fear and angry) from bodily expressions, and agreed with the evidence from neurosciences that the extra striate body area (EBA) is more active for threatening versus neutral expressions and for bodies than faces (Kret, Pichon et al. 2011).

Recognition speed of action tendencies based on the intensity of related emotions

Interesting findings with respect to the action tendency Disappear from View involve that: the shorter perceived users (1/3 or 2/3), the more "guilt" attributed; the longer perceived users (P3/3 or N3/3), the more "shame" attributed. This result simply suggested that users need more time to perceive shame from bodily expressions than guilt. Nevertheless, this finding agreed with the claim in the psychological studies that shame is often a much stronger and more profound emotion than guilt (Barker 2003), though these two emotions are close in the Circumplex model of Russell (Russell 1980). Shame is "a painful emotion caused by consciousness of guilt, shortcoming, or impropriety" and people feel guilty for what they do but feel shame for what they are (Barker 2003). Studies on the perception of dynamics of facial expressions of emotions also observed that users do anticipate emotional dynamics in facial expressions of a virtual character. Subjects may for example perceive the emotion that has being more intense than the one expressed in the last frame of a sequence (Courgeon, Amorin et al. 2010). Similarly, in our experiment, although action tendencies were well recognized in the static pictures (as reported the 1st experimental study), adding a dynamics might have led to some side effects on the perception of intensity of emotions and possibly related emotion categories.

III.6 **Conclusions**

In this chapter, we proposed an approach for designing and evaluating postural expressions of action tendencies. We designed both static and dynamic postural expressions of several action tendencies. These postures were informed by data from the literature and the manual annotations of a video corpus. We used these stimuli to explore the relationships between bodily expressions and action tendencies. We defined experimental methodologies addressing to two perception studies.

In the 1st experimental study, we investigated the spatial configurations of bodily expressions that might convey action tendencies. This study confirmed that bodily expressions might provide information about action tendencies. A decoding study tested 5 pairs of static pictures of postural expressions in terms of action tendencies as well as discrete emotion categories. Subjects reliably recognized the postural expressions of the following action tendencies: Attending, Disappear from View *and* Exuberant. Perception of emotion categories was also consistent with psychological predictions about action tendencies. These results suggested that postural expressions can be useful to express action tendencies in virtual characters.

The 2nd experimental study addressed the research question: to which extent the perception of action tendencies varies as a function of the time that users view the bodily expressions of a virtual character. In order to answer this question, we compared three gate conditions: full time of the animation; third time of the animations, two third time of the animaiton. The results showed that users already recognized action tendencies at the gate of the two third time of the animation. Across overall time duration settings, the action tendency Attending, Disappear from View and In command were better perceived than Antagonistic and Exuberant across overall levels of segmentation and duration of the animation.

In the 1st experimental study, the static posture of Exuberant had been considered as conveying pleasure, pride and joy. However, in this 2nd study concerning the dynamic bodily, subjects attributed different emotional categories to this action tendency: anger, irritation or surprise nonetheless of timing Movement Quality settings and durations of Segmentation and Duration of the Animation. A gap exists between the perception of action tendencies in static postures and that in dynamic bodily expressions. Evidence from neurosciences supported that difference. There might be different neural areas to detect static and dynamic expressions of bodies and faces (Adolphs, Tranel et al. 2003; Kret, Pichon et al. 2011). For example, the superior temporal sulcus (STS) is supposed to be active for dynamic threatening body expressions and to be sensitive to affective information conveyed by the body stimulus. The role of this neural area in detecting affective information from static body expressions is unknown.

This chapter contributed to the analyses of both spatial configurations and recognition speed of bodily expressions that convey action tendencies within a social context in virtual humans. However, this study did not involve human users in interaction and is limit of simulating bodily expressions in virtual environments. In the next chapter, we present one application that establishes a link between human users and virtual characters through simulated bodily expressions. We are interested in whether the simulated bodily expressions in virtual characters would have an effect on the human users during the human-computer interaction.

IV. BODILY EXPRESSIONS IN VIRTUAL CHARACTERS: APPLICATION TO AMBIENT INTERACTION

"*An interaction may be defined as all the interaction which occurs throughout any one occasion when a given set of individuals are in one another's continuous presence.*"

Erving Goffman, The presentation of self in everyday life, 1959

IV.1 **Related work**

Recent years have seen a growing interest in ambient intelligence systems. One challenge of such systems is to discretely and non-intrusively react to where the user is and to what resources are nearby the user in an ambient environment.

As we saw in the previous chapters, virtual characters are expected to support a natural user interaction since they combine intuitive modalities (e.g., gaze, gestures, postures, speech). They are thus potential candidates to enable a discrete and non-intrusive communication with users in ambient environment (Cassell, Sullivan et al. 2000).

Location-aware technologies can be used in ambient environments to provide information about the user's and objects' position inside a room (Steggles and Gschwind 2005).

The combination of virtual characters and location-aware technologies sounds relevant to provide an intuitive assistance to users using nonverbal and spatial behaviors. The main objective of combining those technologies is to create an adaptive and intuitive user interaction in ambient intelligence systems. The assessment of users' reactions to such a location-aware virtual character is thus crucial.

IV.1.1 Virtual characters in ambient environments

Virtual characters have already been integrated in ambient environments. For example, it was observed that a virtual character integrated with a spoken dialogue system provides a pleasant and human-like interaction in an ambient environment, despite its limited contribution to task

effectiveness (Montoro, Haya et al. 2008). A cooking virtual character was also implemented in a kitchen (Miyawaki and Sano 2008). It recognizes the progress of cooking and displays appropriate movies and texts. The pitch of the characters' actions changes as function of the speed of users' movements so as to induce an intuitive interaction. The authors do not mention any evaluation of this cooking character. The EU-founded 2iHome project developed ambient devices to give disabled people greater independence and freedom. A female virtual character displayed on a TV screen in a mocked-up kitchen assists Alzheimer people (Alexandersson 2010).

However, the impact of the integrated virtual characters on users in ambient environments remains unclear.

IV.1.2 Location-aware technologies

Location-aware technologies were used in ambient environments for example to collect positions and movements of the user inside a room (Steggles and Gschwind 2005). RFID sensors were used to record and recognize activities such as preparing breakfast, listening to music, or taking medication. Solutions based on RFID tags were also used at a larger scale in buildings or halls.

Studies were conducted to find the tradeoffs between adaptation to user localization and accuracy (Molina 2010). Image recognition techniques use data from cameras located in an ambient environment. Sequence discovery is one of the machine learning techniques to discover user patterns in ambient environments. For example, the MavHome Project (Rao and Cook 2004) modeled inhabitant actions as states in a Markov model. They used the task-based Markov model to discover high-level inhabitant tasks in unlabeled data. The CASAS smart environment project - the extension of the MavHome Project - used temporal information about entity-tagged data (an entity can be a television, a lamp, etc.).

If combined with location-aware technologies, a virtual character becomes a quite relevant interaction metaphor for ambient environments. For example, it can pro-actively inform users about the technology hidden in the environment and about the services that are available (Kruppa, Spassova et al. 2005). A virtual character can move along the wall and point at physical objects to provide situated assistance in a shopping and navigation application. A virtual character might cheer up users in a cooking navigation system (Miyawaki and Sano 2008).

Nevertheless, studies on the adaptivity of nonverbal spatial behaviors of virtual characters to users' and objects' location remains scarce (Wiendl, Ulhaas et al. 2007).

But embedding a virtual character in an ambient environment raises several questions such as how to select the appropriate spatial behaviors of the virtual character based on those of users? How to evaluate its impact on user's performance and perception?

IV.1.3 Design of ambient systems

Ambient intelligence systems have specific requirements for interaction design since they interact with users when they are not in traditional interaction situations (i.e. sitting at their desk).

Pousman and Stasko (Pousman and Stasko 2006) put on several criteria to evaluate the ambient information systems such as the notification level and the representational fidelity.

The *notification level* refers to the degree to which system alerts are meant to interrupt a user. The ambient display should usually present information without distracting or burdening the user. For example, the Ambient Orb by Ambient Devices is a glass lamp that uses color to show weather forecasts, trends in the market, or the traffics on the homeward commute. This system then has a somewhat low notification level. In contrast, devices that provide alerts by blinking, flashing or opening dialog windows might not be relevant for ambient systems. Thanks to their nonverbal abilities to convey information in a non-intrusive way, virtual characters might keep a low notification level, so that users can be aware of its presence but do not need to always focus on it.

The *representational fidelity* describes a system's display modalities through which the data is encoded, i.e. iconic, symbolic or deictic. Virtual characters may strengthen the representational fidelity by multiplying the encoding modalities that are derived from human communication. The next section discusses these nonverbal spatial behaviors that can be displayed by virtual characters.

IV.1.4 Nonverbal spatial behaviors

We already described nonverbal behaviors in the previous chapters. Here we focus on how several non verbal channels may provide spatial information.

A *deictic gesture* "indicates an object, a location, or a direction, which is discovered by projecting a straight line from the furthest point of the body part that has been extended outward, into the space that extends beyond the person" (Kendon 2004). Deictic gestures might be performed with the hands, the head, body orientation, gaze, and even by protruding the lips, by a movement of the elbow, or with a foot.

In a study of the impact of body orientation on *spatial attention*, Reed et al. (Reed, Garza et al. 2007) presented to human subjects several static images of human bodies representing either action cues (e.g. throwing or running) or static cues (e.g. standing). They measured the response time to targets which appeared on the side corresponding to the direction of the action. Only the body images representing action cues affected the direction of the implied action. This suggests that the orientation of the human body, which can only be found in action cues, improves spatial attention.

Proxemics describes the social aspects of distance between interacting individuals (Harrigan, Rosenthal et al. 2005). Hall proposed a coding system for personal distance (intimate, personal, social, and public) according to different communication functions and the status and affiliation between interactants (Hall 1963). The invasion of the intimate distance zone can elicit strong behavioral and physiological reactions. Several studies argued that this personal space pattern can also be observed in mediated and virtual environments (Kaufman 1974; Reeves 1985; Reeves and Nass 2000; Bailenson, Blascovich et al. 2001; Rehm, André et al. 2005; Vanhala, Surakka et al. 2010). The perceived interpersonal distance between the viewer and mediated character was observed to help individuals to develop parasocial relationships with virtual characters. Furthermore, the reactions evoked by the invasion of the intimate distance zone have been observed to be stronger in the presence of an animated character than in the presence of an inanimate object. People also tend to keep a greater

distance to a mediated person than to a virtual object. People use avoidance and approach to control the proximity to others. Avoidance movements are usually performed in the context of aversive or problematic conditions that require an enhanced control to ward off negative consequences (Smith 2008). Variations of personal space management can be related to several factors such as gender, age, race, status, degree of acquaintance … (Lohse 2009).

The MirrorSpace uses proxemics in an ambient awareness display for video communication (Roussel, Evans et al. 2004). It uses physical proximity to control the blurring of the displayed image. The image becomes more clear as the user gets closer to it. The image becomes blurry when the user moves away. The assumption is that the closer the user stands to the device, the more engagement the user wants from the interlocutor on the other side.

To conclude, virtual characters are able to use several nonverbal behaviors that might convey spatial information (deictic gestures, proxemics), thus affording multiple levels of representational fidelity.

IV.1.5 Evaluation of virtual characters: social presence, adaptation and engagement

Social presence (Blascovich 2002) is different from physical presence that describes a subjective feeling of being in one place or environments. Several studies showed that virtual characters might create a feeling of social presence (Bailenson, Blascovich et al. 2001; Blascovich 2002; McQuiggan, Rowe et al. 2008). It is "the illusion that the sensory stimuli created by technological devices are real" (Kasap and Magnenat-Thalmann 2009).

Adaptation to the user is considered as one of the required capabilities for the success of a virtual character (Ruttkay and Pelachaud 2004). A virtual character should adapt its own behavior to the static or dynamic characteristics of the user. We define perceived adaptivity as the perception of users about how, when and what the virtual character reacts accordingly (e.g. by expressing nonverbal spatial behaviors).

Users should also have the feeling of control over human-computer communications (Spiekermann 2007). Evaluation studies thus should evaluate if users perceive the virtual character notification system useful and easy (Davis 1989; Venkatesh and Davis 2000).

During human interactions, people start, maintain and end a perceived connection to each other. This process has been defined as *engagement* by psycholinguistics (Sidner and Lee 2007). Engagement might occur during human-computer interaction, in which people would have interest in the content or activity of an experience, regardless of the interactive medium (Dow, Mehta et al. 2007). From the psychological perspectives, the process of engagement refers to a state in which subjects' energy and attention are focused on a coherent set of stimuli or meaningfully related activity and events (Witmer and Singer 1998). Similar terms such as deep play, holding power and magic circle are used to describe the level of user engagement in video games (Dow, Mehta et al. 2007). Measures of engagement in the field of virtual characters rely on observable behaviors such as gaze (Poggi, Pelachaud et al. 2000; Peters, Pelachaud et al. 2005; Nakano and Ishii 2011). In face-to-face human conversations, people continuously confirm if interlocutors are engaged in the current conversation based on their gaze. Interlocutors who intent to disengage from a conversation will start to avert their gaze (Peters, Pelachaud et al. 2005). During social interaction, the overt behaviors such as gaze and eye contact may signal user's intention to become or remain engaged (Poggi, Pelachaud et al. 2000). Gaze provides critical cues to the focus of attention, where engagement can be created and maintained (Sidner and Lee 2007). Gaze can be measured in

terms of frequency (number of glances at a partner), total duration or total gaze (total number of seconds looking at partner), proportion of time looking during a specified activity (e.g. listening and speaking), average duration (mean duration of individual glances) and standard deviation of glances (Harrigan, Rosenthal et al. 2005).

Several studies investigated the impact of technologies on presence and engagement. Dow et al. (Dow, Mehta et al. 2007) explored the effect of immersion on presence and engagement in a qualitative study. Three experimental conditions of an interactive drama were compared: augmented reality, speech-based desktop interaction and keyboard-based desktop interaction. They found that increased feeling of presence did not necessarily lead to more engagement. Only qualitative studies such as observations and interviews were used for investigating these relationships between interactive contexts, presence and engagement. McQuiggan et al. (McQuiggan, Rowe et al. 2008) studied the effect of empathetic characters on perceptive presence. Students interacted with narrative-centered learning environments. The results showed that the use of empathetic characters had a positive impact on perceived presence, involvement and control in learning situations for the two groups. The measures of involvement and control were performed using a questionnaire of perception (e.g. "how much did the visual aspects of the environment involve you?", "how involved were you in the virtual environment experience?", "Were you involved in the experimental task to the extent that you lost track of time?"). However, these measures involved only perceived involvement/control but not overt behaviors.

The above described studies modeled the direct effect of the technologies on presence or engagement. However, simple two-variable (predictor-outcome) models of interaction are not enough to explain how the interaction occurs. These models do not consider the mediating role of perceptual factors and the possible moderating role of individual differences.

Interaction between communication technology and users, or between users through technology, is derived from technological attributes that enable reciprocal communication or information exchanges in mediated contexts. These technological attributes are defined as interactivity by Bucy and Tao (Bucy and Tao 2007). Human-Computer Interaction covers three approaches to modeling interactivity: 1) the *message-centered approach*, which considers the impact of media stimuli on user perceptions, 2) the *structural approach* which considers the impact of interactive attributes on media effects such as affect, behavior and cognition, and 3) the *perceptual approach*, which considers the impact of perceived interactivity on media effects. However, none of the interactivity models explains how the relationship between media stimuli and effects can be moderated by different user characteristics and mediated by different psychological states. There is a lack of considering mediating or moderating relationships in interactivity modeling. Mediating relationship attempts to identify a variable or variables through which the independent variable acts to influence the dependent variable. Moderating relationships refer to situations in which the relationship between independent and depend variables changes as a function of the level of a third variable (Baron and Kenny 1986). Bucy and Tao (Bucy and Tao 2007) proposed to incorporate interactive attributes, user perceptions, individual differences, and media effects measures for a systematical examination of the definition, process and consequences of interactivity on users. There may be one or more variables mediating or moderating the relationship between interactivity and its outcome on users. For example, the relationship between interactivity and performance is moderated by the level of web experience (Bucy and Tao 2007); the effect of exposing a pre-game story video on people's evaluation of a game was observed to be mediated by the feeling of presence during the game (Park, Lee et al. 2010).

IV.2 Designing and evaluating a location-aware virtual character in Ambient Interaction

In this section, we explain how we designed a location-aware virtual character that adapts its spatial behavior to users and to objects locations during a search task in a smart room.

The next section will describe an experimental evaluation comparing this adaptive virtual character with another virtual character that does not perceive nor use the location of users and objects.

IV.2.1 System

We used an intelligent room, called "iRoom" (Bellik, Reba ïet al. 2009), that was designed for conducting evaluation studies of interaction techniques in an ambient intelligent environment. The room contains sensors that allow for context-awareness capabilities (Figure 1).

The walls of the iRoom are equipped with four Ubisense© sensors in order to detect the location of users and objects in the room. Acquiring location information in open spaces has been studied from the beginning of ubiquitous computing, resulting in various models and techniques (Davis 1989; Eissfeller, Gaensch et al. 2004). The Ubisense© system relies on a network of sensors and a set of tags that can be attached to objects or to users. Contrary to other solutions using video processing, this location system is robust to occultation: the tags do not need to be visible or to be detected. This feature can be useful for example for searching tasks, in which objects can be hidden in drawers. The average precision is 20cm, though it is variable depending on the location. The system can track objects locations and detect if an object has entered a certain zone or collided with another object.

We integrated the virtual character platform MARC (Courgeon, Martin et al. 2008) with the smart room environment (Figure 40). As we described in the previous chapter, this virtual character platform includes editors for specifying body animations but also facial expressions. These behaviors can be blended and interrupted during real-time rendering which makes this platform quite original and relevant for conducting evaluation studies about interaction.

A component was developed. It receives the location events sent by the Ubisense sensors, analyzes them and generates commands that are sent to the MARC virtual character in order to generate appropriate spatial nonverbal behaviors.

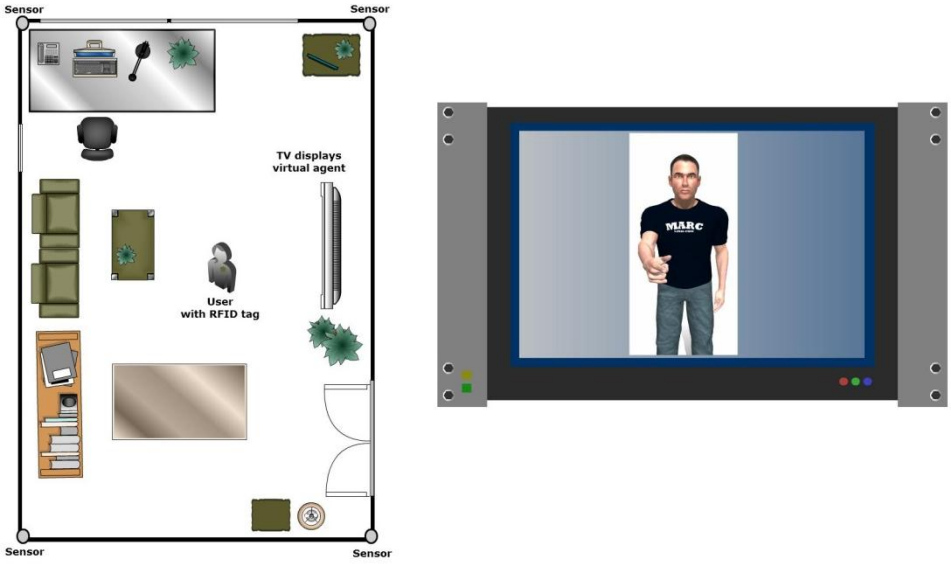


Figure 40: The ambient intelligent room used in the experiment with Ubisense location sensors at the four corners of the room (left); a MARC virtual character is displayed on the TV display (right)

IV.2.2 Hypotheses

We compared two conditions in a between-subject design. In the adaptive condition, the virtual character uses its knowledge about locations of objects and users to select nonverbal and spatial behaviors. The adaptive virtual character was thus able to look at users, point at objects, look at objects, orient the body towards objects, and approach in the direction of some specific objects. These different nonverbal spatial behaviors were used incrementally and progressively when users appeared to have difficulties in executing the task during the experimentation session. In the control condition, a non-adaptive virtual character provided information only via speech and facial expressions (Table 56).

Based on the studies described in the related work section, we have the following hypotheses:

H1: Users in the adaptive condition will report a higher perceived social presence than users in the non-adaptive condition.

H2: Users in the adaptive condition will report a higher perceived adaptivity than users in the non-adaptive condition.

H3: Users in the adaptive condition will perform difficult tasks more quickly than users in the non-adaptive condition.

H4: Users in the adaptive condition will be more engaged with the virtual character than users in the non-adaptive condition.

IV.2.3 Scenario

We aimed at investigating the impact of the adaptive capacities of our virtual character on the user. We selected a search task in which the user has to move around in the ambient room to find objects. Due to the nature of this task, we expected rich and continuous location

information and behaviors. Furthermore, we expected that this task would keep the user in continuous interaction with the virtual character, i.e. that the user would have to look frequently at the virtual character.

We selected a procedural searching task. Unlike conversation or dialogue tasks, procedural tasks are strictly defined so that each step is clear and unambiguous to users. A procedural task breaks down into sequential steps, which makes user interaction arranged sequentially.

The interactions are mainly based on nonverbal cues; the cases in which the virtual character uses speech output is for informing the user of the next target object to search for and pointing to objects (in the adaptive virtual character condition).

Six target objects were equipped with Ubisense© tags so that the system knew how far the user was from each of these objects. The objects that we selected were: an adapter, an envelope, a jacket, a DVD, a medicine box and a folder. Two objects were hidden in drawers (the medicine box and the DVD). The four other objects were visible: the adapter was on the desk, the envelope was on the coffee table, the folder was on the bookshelf, and the jacket was on clothes stand. The initial position of each object was the same for all users. The position of these target objects were chosen so as to take advantage of the whole space and require the user to move across the entire room. Three distractor objects (objects that look similar to target objects) were also equipped with Ubisense© tags to detect when the user was taking a wrong object. Task difficulty for searching for a target object was estimated by considering the number of distractor objects. Searching a target object that has distractors was expected to be more difficult than searching a target object that has no distractors. Users performed tasks in the same order of increasing difficulty in order to have the same learning effect for each of the participants.

The virtual character was displayed on a flat 42-inch TV set. The head movements, gestures, body orientation and gaze of our virtual character could be directed towards a specific location in the room such as the user's location or objects' location.

Four communicative acts were selected as being relevant to the scenario: Point, Confirm, Disapprove and Congratulate. For each of them, several postural expressions were designed based on a small video corpus that we collected previously.

No background was displayed behind the virtual character to avoid distracting the user from the behavior of the virtual character. An animation was also designed to have the virtual character walk, simulating an "approach" behavior towards a location or an object in the real world.

Table 56: The behaviors of the virtual character in the two experimental conditions: non-adaptive character vs. adaptive virtual character.

BEHAVIORS		CONDITIONS	
		NON-ADAPTIVE VIRTUAL CHARACTER	ADAPTIVE VIRTUAL CHARACTER
Spatial behaviors	Proximity (moving towards objects or users)	NO	YES
	Orientation (head and gaze turn towards objects or users, point)		
Postures	Communicative postures (disapprove, congratulate, confirm)	NO	YES
Facial expressions	Positive and negative emotional facial expressions depending on communicative act	YES	YES

A test was conducted with eleven subjects rating these postural animations in terms of valence (positive vs. negative) and communicative acts. The postural animation that best expressed each communicative act was kept for the user study.



Figure 41: Some illustrations of postures designed and validated for the following communicative acts (from left to right): point, disapprove, congratulate

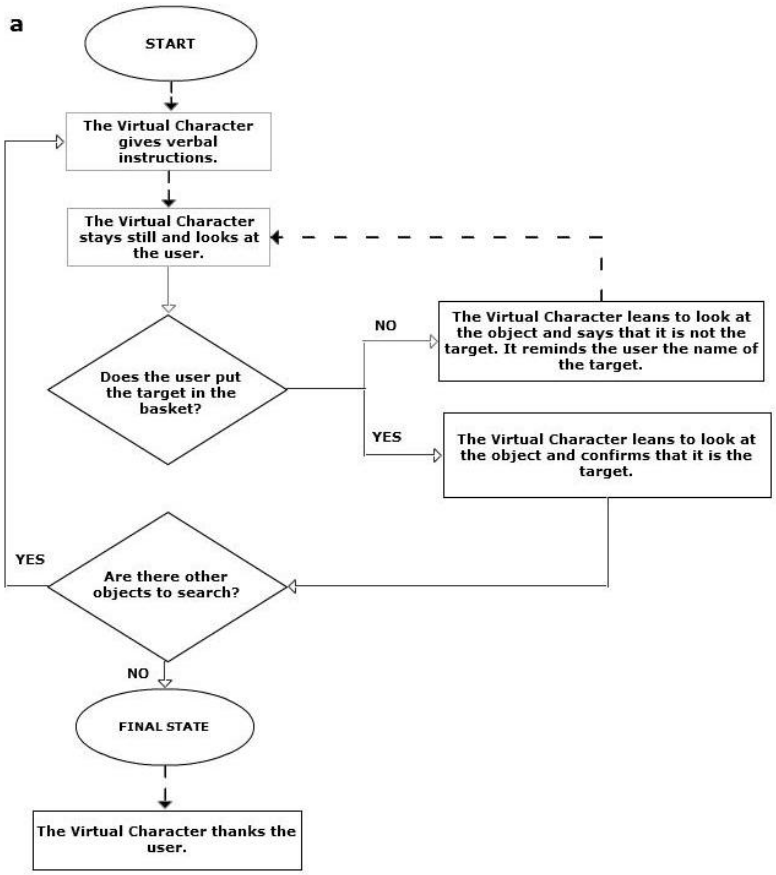
IV.2.4 Procedure

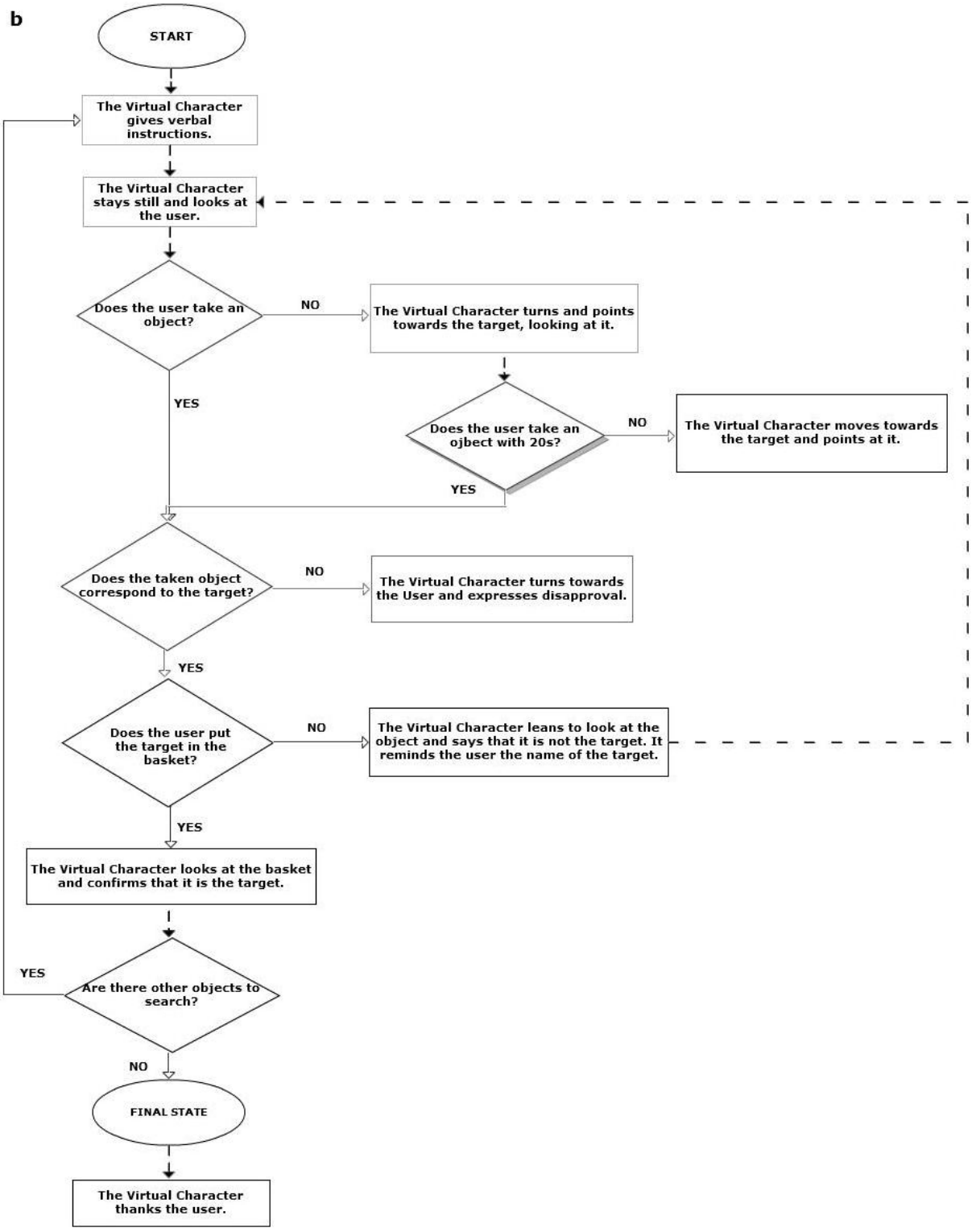
Before starting the session, users were briefed about the search task. There was a time constraint since the background story was about a friend stuck at an airport and asking the user to collect some objects from his flat. The user received an Ubisense tag to attach to her clothes. At the beginning of the session, the virtual character asked the user to find the first target object and to put it into the basket in front the TV set.

IV.2.5 Modeling the ambient interaction with the virtual character

The interaction between the user and the virtual character was specified using a graph (Figure 42).

Figure 42: The automata managing the interaction between the user, the objects, the ambient system and the virtual character: a) interacting with the non-adaptive virtual character b) interacting with the adaptive virtual character.





IV.2.6 Data collection

We collect data about user characteristics, users' behavior and perception using methods such as sensing users' location during the session, structured questionnaire, and interviewing. The behaviors of the users were manually annotated (based on video recordings) including gaze,

movements and actions regarding the objects and the task. The level of user engagement was estimated from users' gaze behavior. We examined user perception in terms of perceived social presence, perceived adaptivity, usability and satisfaction. Our study attempts to clarify the role of an adaptive virtual character on users' engagement by considering the mediating of perceived presence and the moderating role of user characteristics.

IV.2.7 Participants

Thirty-two participants completed the experiment: 16 female, 16 male; aged 20-58, average 30.5; 75% European, 18% African, 3% American and 3% Asian.

40.6% of participants had little or no interaction with a virtual character before the session.

59.3% of participants had not heard of ambient intelligence.

All users found the same six objects in the same order.

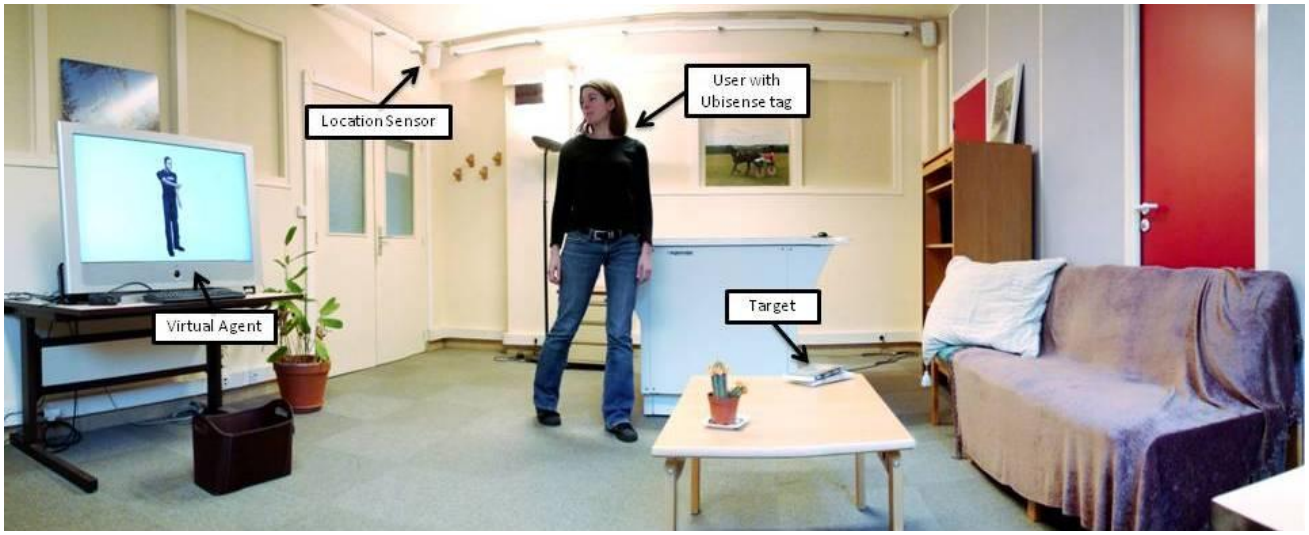


Figure 43: The ambient intelligent room with the virtual character displayed on a TV set, a user equipped with a Ubisense tag, a target object on the table, and a location sensor on the ceiling.

IV.2.8 Data collection

Four types of data were collected (Table 57).

The files logged by the Ubisense detection system for the user's tag and for the object's tags were used to compute the duration of the task.

The video recording of the sessions provide data on user's non-verbal behavior (e.g. gaze and body orientation towards the virtual character, moving patterns in the room). As discussed earlier, gaze signals engagement during human-human communication. We thus measured the user's gaze (in term of frequency and time length) to evaluate user engagement in the interaction with the virtual character.

We designed a post-hoc questionnaire for evaluating how users perceived the social presence and the adaptivity of the virtual character.

Finally, semi-structured interviews were used.

Table 57: Collected data.

	User's Location	User's Actions	User Perception	
Technique	Location Sensoring	Video	Questionnaire	Interview
Equipment	Ubisense	Camera	Web-based Questionnaire	Camera
Objective	Performance of individual search tasks	Manual annotation and automatic analyzes of user's behaviors (look, move and take actions) including engagement measures	To evaluate perceived social presence and perceived adaptivity that the virtual character made to the user in their shared environment.	To shape the user experiences (identify anomalies) regarding the user-character interaction issues.
Nature of Analysis	Quantitative	Quantitative	Quantitative	Qualitative
Related Hypothesis	H3	H4	H1, H2	/

IV.2.8.1 Video annotation

We collected 31 videos (the video recorded for one user was not recorded due to technical problems). The total length of the video corpus is 4 hours and 15 minutes.

We selected three frequent actions that we observed while previewing the videos and that are relevant for testing our hypotheses: look, move and take (Table 58).

Table 58: The coding scheme defined for guiding the manual annotation of the collected videos of user's behaviors

	Actions
Look	Look around
	Look at an object
	Look at the virtual character for a following instruction
	Look at the virtual character for other reasons
Move	Move towards the target object (MTT)
	Move towards elsewhere than the target object (MTEW)
	Move towards the virtual character with an object (MTVO)
	Move away from the virtual character with an object (MAAO)
	Stand still waiting for validation from the virtual character (SSV)
	Take the target (TT)

Take	Take a distractor (TD) Show a taken object (STO)

The video were annotated using the Anvil tool (Kipp 2003). First, the sequential tasks were identified and the videos were segmented accordingly. Then, user's actions were annotated using the coding scheme described in Table 1.

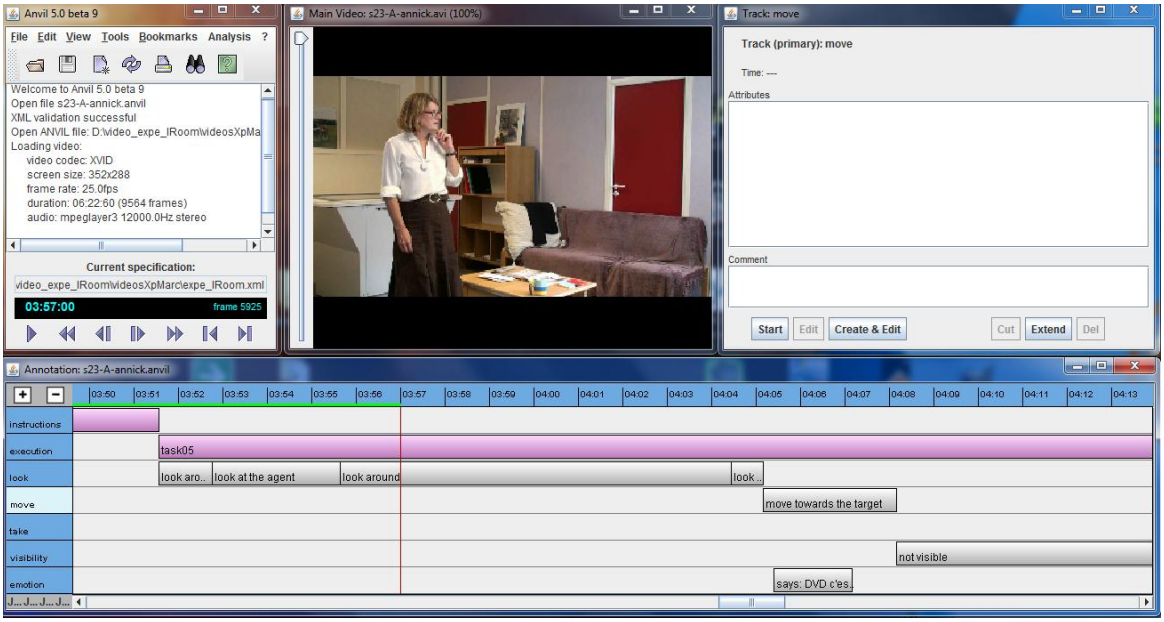


Figure 44: Annotation of the behaviors of a user with the Anvil tool (Kipp, 2003)

IV.2.8.2 Post-hoc questionnaire

We asked users to answer a five-point Likert scale questionnaire (from 0 strongly disagree to 4 strongly agree) representing the following dimensions: presence, adaptivity, usability, and satisfaction.

The first part of the questionnaire aimed at measuring the perceived presence of the virtual character using 4 items adapted from (Venkatesh and Davis 2000): “I perceive that I am in the presence of another person in the room with me.”, “I feel that the person is watching me and is aware of my presence”, “The thought that the person is not a real person often crosses my mind”, “The person appears to be sentient (conscious and alive)”.

The second part of the questionnaire consisted of five items evaluating how users perceived the adaptivity of the virtual character: “The activities of the virtual character meet my needs”, “The virtual character expresses emotions based on what I do”, “The virtual character adapts its gaze my actions”, “The virtual character adapts its position to my actions”, and “The virtual character adapts its bodily expressions to my actions”.

The third part of the questionnaire aimed at measuring usability. It included the following items (adapted from the UTAUT questionnaire (Venkatesh and Davis 2000): “My interaction with the virtual character is clear and understandable”, “Learning to interact with the virtual character is easy for me”, “Using the virtual character takes too much time from my normal home activities”, “It takes too long to learn how to use the virtual character to make it worth the effort ®”, “Using the virtual character would prevent me accomplishing my tasks more quickly”, “If I use the virtual character, I will spend less time on routine home tasks”, “I lose control over using the virtual character”, “I think that using the virtual character fits well with the way I like to work”.

The fourth part of the questionnaire considers likeability and perceived helpfulness (adapted from the UTAUT questionnaire (Venkatesh and Davis 2000)): “I think that the avatar did not help me a lot for my task ®”, “I find that the avatar is not adaptive in the way I supposed ®”, “The virtual character makes my work more interesting”, “Working with the virtual character is fun” and “The virtual character is okay for some jobs, but not for the kind of job that I want ®”, “The virtual character gave me a warm impression”.

IV.2.9 Quantitative results

IV.2.9.1 Validation of measures

Cronbachs alpha indicates the degree which a set of questions measures a single one-dimensional latent construct. We used it to validate the questionnaire. We computed a value of 0.80 for the section about presence, 0.85 for the section about adaptivity, 0.73 for usability, and 0.73 for satisfaction. The American Psychological Association considers a questionnaire as reliable when the alpha coefficient is above 0.7 ($0< \alpha \leq 1$).

IV.2.9.2 Hypothesis tests

We performed one-tailed Students t tests to compare the two conditions on interaction and post-interaction measures. We assumed that the two samples came from normal distributions with unknown and possibly unequal variances. We used Satterthwaites approximation for the effective degrees of freedom.

IV.2.9.2.1 Social presence

H1: Users in the adaptive condition will report a higher perceived social presence than users in the non-adaptive condition.

Results: We observed a significant difference in the answers one item of the social presence section. Unlike users interacting with the non-adaptive virtual character, users interacting with the adaptive virtual character felt that the virtual character was watching them and was aware of their presence ($t(16)=3.007$, $p= 0.003$).

IV.2.9.2.2 Perceived adaptivity

H2: Users in the adaptive condition will report a higher perceived adaptivity than users in the non-adaptive condition.

Results: The adaptivity of the virtual character was perceived as significantly higher by users in the adaptive condition than those in the non-adaptive condition ($t(16)=2.046$, $p=0.025$). Users interacting with the adaptive virtual character felt that the virtual character reacted to their needs ($t(16)=1.781$, $p=0.043$), and that, depending on the users' actions, the virtual character expressed emotions ($t(16)=2.851$, $p=0.004$), gaze ($t(16)=3.128$, $p=0.002$), location ($t(16)=1.980$, $p=0.029$) and bodily orientation ($t(16)=2.074$, $p=0.023$).

IV.2.9.2.3 Task performance

H3: Users in the adaptive condition will perform difficult tasks more quickly than users in the non-adaptive condition.

Results: We did not observe any difference in the duration of the whole session across the two conditions ($t(16)=0.044$, $p=0.483$). Yet, we did observe an effect related to task difficulty. Users interacting with the non-adaptive virtual character took more time to complete difficult tasks (Task 4, 5, 6) than easy tasks (Task 1, 2, 3). They performed Task 2 significantly faster than Tasks 4, 5 and 6 ($t(16)=-2.378$, $p=0.016$; $t(16)=-2.595$, $p=0.009$; $t(16)=-5.082$, $p=0.0001$) compared to those interacting with the adaptive virtual character ($t(16)=-4.821$, $p=0.209$; $t(16)=-1.642$, $p=0.060$; $t(16)=-2.918$, $p=0.005$). Task 3 was performed significantly faster than Tasks 4, 5 and 6 ($t(16)=-2.3412$, $p=0.017$; $t(16)=-2.086$, $p=0.024$; $t(16)=-4.948$, $p=0.0001$) compared to the condition of the adaptive virtual character ($t(16)=-1.333$, $p=0.098$; $t(16)=-0.587$, $p=0.280$; $t(16)=-2.563$, $p=0.009$). Task 1 was performed faster than Task 4 and Task 6 ($t(16)=-2.009$, $p=0.031$; $t(16)=-3.99$, $p=0.0004$), but not Task 5 compared to the condition of the adaptive virtual character ($t(16)=-4.528$, $p=0.001$; $t(16)=-1.522$, $p=0.073$).

IV.2.9.2.4 Engagement

H4: Users in the adaptive condition will be more engaged with the virtual character than users in the non-adaptive condition.

We estimated users' engagement via gaze shift frequency and the duration of gaze. We examine the data from the manual annotations of videos with respect to the gaze-related action units. Each behavior unit contains two types of data: frequency and length. An unpaired sample and unequal sample size statistic test was used to compare the two groups. Since the test is unpaired, the fisher-Snedecor F-test was performed to assess the equality of

variance. If variances were equal, Students T-test was performed; if variances were not equal, Satterthwaites approximate t test was performed.

When presented with an adaptive virtual character, the user showed a higher level of visual attention towards the virtual character in terms of action frequency (mean difference = 145 shifts) and action length (mean difference = 351.6 ms). T tests of the difference between means produced a statistically significant result $t(29) = 4.9079$, $p = .0003$, $t(29) = 2.95178$, $p = .0062$), $\alpha = .05$.

Hypothesis H4 is supported: the users in the adaptive virtual character group paid a higher level of visual attention towards the virtual character. One possible interpretation is that users in the adaptive condition were more engaged in the interaction with the virtual character.

The absence of adaptivity in the virtual character let the user to make more attention shifts and attention deployments towards the shared space in terms of action frequency (the mean difference = 61 shifts) and action length (the mean difference = 715.6 ms). The results were respectively: $t(29) = 2.052$, $p = .052$; $t(22) = 2.179$, $p = .038$, significant at $p < .05$.

Table 59: T-test results: comparing looking behavior of users in the two groups (the values of significance are marked in grey shading)

		look at the virtual character		look at the virtual character for a following instruction		look around		look at an object	
		Frequency (units)	Length (s)	Frequency (units)	Length (s)	Frequency (units)	Length (s)	Frequency (units)	Length (s)
Equal Variances assumed	F	2,883	1,69536	4,2	3,99065	3,78671	2,38598	1,1414	2,3743
	p-value (2-tailed)	0,05	0,32178	0,01054	0,0134	0,01706	0,11215	0,80957	0,10801
	Equal Variance	YES	YES	NO	NO	NO	YES	YES	YES
	Mean difference	10,14	24,60	1,08	3,33	3,22	40,97	1,72	33,91
Equal Variances not assumed	t	4,9079	2,9528	2,609	2,382	2,052	2,179	1,474	2,130
	df	29	29	22,0619	22,3775	22,7101	29	29	29
	p-value (2-tailed)	0,00003	0,006	0,016	0,026	0,052	0,038	0,151	0,041
	Mean difference	10,14	24,60	1,08	3,33	3,22	40,97	1,72	33,91

IV.2.9.3 Analysis of user data

IV.2.9.3.1 Correlations between the possible factors

To explore the relationships between user actions and user perception, we computed Pearson's correlation coefficient between:

- user action (look at the virtual character, look at the virtual character for a following instruction, look around, look at an object, move elsewhere than in the direction of the target, move towards the target)
- user perception (perceived ease of use, perceived efficiency, perceived control, attitude, affect towards use, likeability, presence, adaptivity)
- 28 items from the questionnaire of user perception (5-point likert scale)
- gender of the subjects (Male, Female)
- prior expertise on virtual characters and Ambient Intelligence (5-point likert scale)

We present p-values for Pearson's correlation to represent the correlations between the dimensions of user action and the items from the questionnaire of user perception. The Pearson's p-value is the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero. The correlation is significant, if $p < .05$. As shown in Table 60, twelve items out of the total of twenty-eight were correlated with the user's actions.

Table 60: Correlations between user's behavior and items from the perception questionnaire (the significant correlations are marked with an asterisk) (FQ: frequency, DR: duration)

DIMENSIONS		PERCEIVED EASE OF USE	PERCEIVED EFFICIENCY	PERCEIVED CONTROL	ATTITUDE	AFFECT TOWARDS USE	HELPFULNESS	AFFECT TOWARDS USE	LIKEABILITY	SOCIAL PRESENCE	SOCIAL PRESENCE	ADAPTIVITY	ADAPTIVITY
items		My interaction with the virtual character is clear and understandable.	Using the virtual character would prevent me accomplishing my tasks more quickly. ®	I think that using the virtual character fits well with the way I like to work.	Using the virtual character is a bad idea. ®	I find the interaction with a virtual character to be enjoyable.	I find that the avatar does not help me a lot in my task execution. ®	Working with the virtual character is fun.	I find that the virtual character gives me a warm impression.	I feel that a person is watching me and is aware of my presence.	The thought that the person is not a real person crosses my mind often.®	The activities of the avatar meet my needs	The avatar expresses emotions based on what I do
look at the virtual character	FQ	0,07	0,66	0,29	0,37	0,18	0,02*	0,56	0,17	0,02*	0,20	0,51	0,01*
	DR	0,01*	0,35	0,45	0,52	0,05	0,06	0,29	0,23	0,03*	0,18	0,32	0,03*
look at the virtual character for a following instruction	FQ	0,65	0,36	0,21	0,14	0,61	0,83	0,17	0,85	0,92	0,57	0,29	0,51
	DR	0,72	0,18	0,26	0,02*	0,64	0,96	0,02*	0,59	0,77	0,94	0,43	0,83
look around	FQ	0,96	0,06	0,01*	0,27	0,37	0,36	0,92	0,19	0,38	0,02*	0,01*	0,24
	DR	0,74	0,02*	0,00*	0,51	0,03*	0,67	0,86	0,07	0,33	0,01*	0,08	0,62
look at an object	FQ	0,49	0,89	0,24	0,09	0,76	0,09	0,75	0,14	0,53	0,15	0,02*	0,59
	DR	0,58	0,80	0,41	0,35	0,50	0,06	0,93	0,03	0,13	0,01*	0,33	0,33
MTEW	FQ	0,38	0,41	0,15	0,86	0,81	0,50	0,80	0,82	0,31	0,22	0,00*	0,06
	DR	0,97	0,44	0,05	0,39	0,82	0,78	0,85	0,71	0,38	0,25	0,01*	0,11
MTT	FQ	0,65	0,15	0,08	0,09	0,32	0,41	0,16	0,25	0,95	0,08	0,08	0,43
	DR	0,30	0,14	0,06	0,07	0,06	0,02*	0,16	0,12	0,46	0,02*	0,11	0,10

Users who found that the interaction with the virtual character was clear and understandable are those who spent considerable time to look at the virtual character during the task (p=.01). Users who found that using the virtual character would not prevent them accomplishing the task more quickly are those who spent a lot of time to look around in the room (p=.02). Users who thought that using the virtual character fits well with the way they like to work are those who had frequently looked around (p=.01) and who had took significant time to look around during searching (p=0).

Users who found that using the virtual character in ambient is a good idea, and found that working with the ambient virtual character is fun spent a lot of time to look at the virtual character during the inter-task phase (p=.02, p=.02). However, users who found the interaction with the virtual character to be enjoyable are those who had took considerable time to look around for searching (p=.03). Users who found that the virtual character helped them a lot in the task execution are those who had frequently looked at the virtual character during the task execution and had took time to move towards the target (p=.02). Users who found that the virtual character did give them a warm impression are those who had spent time to gaze at a specific object (p=.03).

Users who felt that the virtual character was watching them and was aware of their presence are those who had frequently looked at the virtual character and took considerable time to gaze at the virtual character during task execution (p=.02, p=.03). However, users who confirmed that the thought that the virtual character is a real person often crossed their mind are those who had frequently looked around, looked at some specific objects, and moved towards the target (p=.02, p=.01, p=.01, p=.02).

Users who found that the virtual character had expressed emotions based on what they were doing are also those who had frequently looked at the virtual character and spent a lot of time to gaze at the virtual character (p=.01, p=.03). Users who found that the activities of the avatar met their needs were those who frequently looked around, moved around, looked at a specific object, and spent a lot of time moving around without going in the direction of the target.

As shown previously, no significant difference was found between the group Adaptive and the Control group in terms of task performance. However, Table 61 shows that users who perceived a high level of efficiency are those who had frequently looked around and taken considerable time to look around. Users who agreed with a high level of adaptivity of the virtual character are those who had frequently shifted gaze towards the virtual character. This correlation corresponds to the facts that the group Adaptive looked more at the virtual character than the group Non-Adaptive, and the group Adaptive perceived a high level of adaptivity in the virtual character than the group Non-Adaptive.

IV.2.9.3.2 Structural modeling

As discussed previously, simple two-variable models of interaction are not enough to explain complex relationships among several independent and dependant variables that co-occur during the interaction. We used Structural Equation Modeling (SEM) to test and estimate the causal effects of the experimental conditions (adaptive virtual character vs. non adaptive) and the user characteristics regarding perceptual factors and user engagement.

We did not use classic multivariate modeling techniques (i.e. ANOVA), as they only examine direct relationships between independent and dependant variables. We built a structural model using path analysis that could convey causal assumptions based on the maximum

likelihood estimation (Figure 45). Path analysis was chosen because this type of structural model examines the relationships among measured variables. It shows statistically the effects of independent variables on a particular dependant variable. We used the multiple regression approach to the structural equation modeling.

In regression models, the dependent variables regress on the independent variables. The dependent variables are predicted by the independent variables. Our model contains two dependant variables (perceived social presence and user engagement) that are predicted by three independent variables (adaptive virtual character vs. control, expertise and gender).

Regression analysis helps to quantify the importance of the relationship between different variables. We used Standardized Regression Weights (β) to measure the importance. User engagement is predicted by the experimental conditions adaptive agent vs. control (β= .583), as well as by the gender (β= .159) and the user expertise (β= .202). These three variables explain 46% of variance of user engagement.

We performed a goodness-of-fit test to calculate how similar the predicted data are to matrices containing the relationships in the actual data. We used Pearson's chi-square (X^2) as goodness-of-fit index because it is the fundamental measure of fit considering both the sample size and the difference between the observed data and the hypothesized model. The X^2 will give a non significant p value if the model fits the data. Applying the fit test with our model showed that the model fits the data very well, $X^2 = 2.778$ (DF=5), p =. 734.

Table 61: Significant correlations between independent variables and dependant variables

	Perceived social presence	User engagement
Adaptive virtual character vs. Control	0.0821	1
Gender	0.0524	-0.056
Expertise	0.1626	-0.0188

Standard regression models assume that independent variables are measured without errors. As shown in Figure 45, the model involved two unobserved independent variables that are supposed to be error1 and error2, which explain the unknown causes affecting perceived social presence and user engagement, respectively.

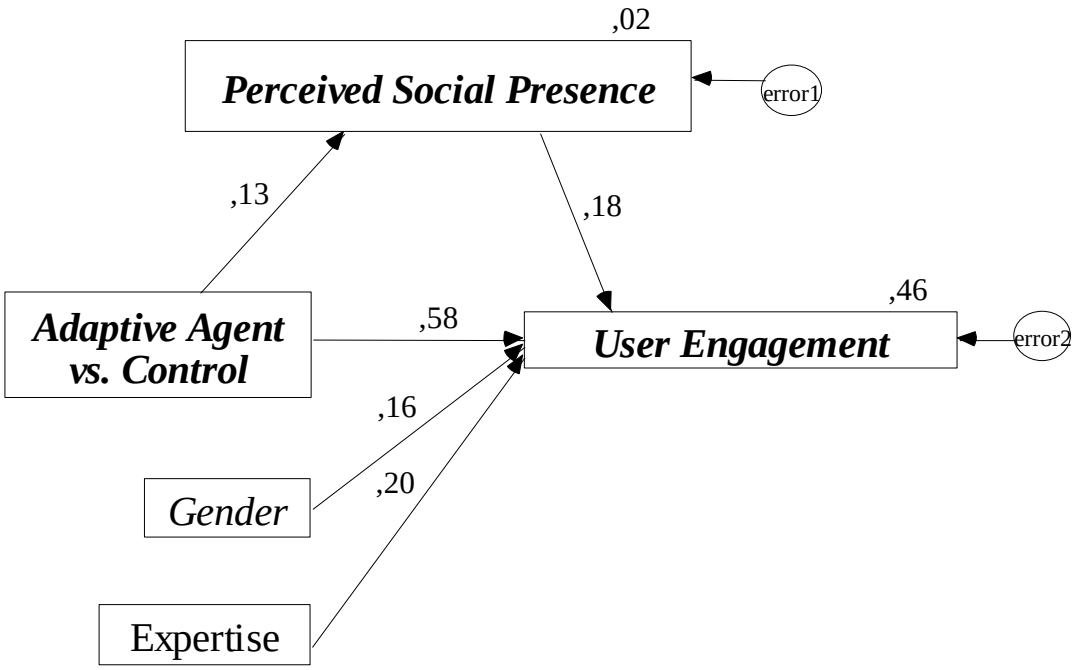


Figure 45: Structural model explaining relationships between the conditions of virtual character, perceived social presence, user characteristics and user engagement

IV.2.9.3.2.1 Mediating relationships

In the above regression model, the independent variables are modeled as having indirect effects through a mediator on the dependent variables. We investigate the indirect effect of the adaptive virtual character on the user engagement through the mediation of perceived presence. Baron and Kenny (Baron and Kenny 1986) defined a set of prerequisites to claim a mediating relationship:

- Significant correlations between the predictor and outcome
- Significant correlations between the predictor and the mediator
- Significant correlations between the mediator and the outcome
- Reduction in the effect of the predictor on the outcome (indirect effect)

The amount of mediation, called the indirect effect, is defined as the reduction of the effect of the predictor on the outcome (Baron and Kenny 1986). The standardized direct effect of adaptive virtual character vs. control on user engagement is .583. The standardized indirect effect of adaptive virtual character vs. control on user engagement is defined as the product of two standardized direct effects: the standardized direct effect of adaptive virtual character vs. control on perceived social presence (0.13) and the standardized direct effect of perceived social presence on user engagement (0.18). The product of these two standardized direct effects is $0.13 \times 0.18 = .023$.

Two methods are advocated by researchers to test significance of indirect effects (Preacher and Hayes 2008): the Sobel test and bootstrapping. We choose bootstrapping because it does not impose the assumption of normality on the sampling distribution. In addition, the Sobel test requires large samplings. The bootstrapping results indicated that perceived social presence is a significant mediator at the confidence interval of 95%.

-.491 is the mean of the indirect effect estimate calculated across all bootstrap samples (1000). While the lower confidence interval is -2.512, the upper confidence interval is 0.956.

IV.2.9.3.2.2 Moderating relationships

When measuring the effect of the adaptive virtual character vs. control on user engagement, we found that this effect is not strong enough as we expected. We assumed that individual differences may lead to different user reactions to the virtual character, adaptive or not. In the post-experimental phrase, we collected user information regarding their gender, age and user expertise on ambient intelligence and virtual characters. We assumed that these variables might moderate the relationships between the experimental conditions and user engagement.

However, as shown in Table 61, we only consider variables that have significant correlations with the dependant variables (i.e. perceived social presence and user engagement): gender and user expertise. The variable age was not correlated with any other variables and was removed from the model.

We performed a significant test of moderation with respect to gender and user expertise. Graphing moderation (for both categorical and continuous moderators) uses three values as a common practice in Behavior Sciences (Frazier, Tix et al. 2004): high (1 SD above the mean), medium (the mean) and low (1 SD below the mean). We dummied the codes for the conditions (1: adaptive virtual character group; 2: control group) and Gender (1: Female; 0: Male). The user expertise consists of two parts: 1) Virtual Character Experiences: prior experiences on virtual characters, 2) Ambient Experiences: prior experiences on ambient environments.

Table 62 : Significance of slopes for moderation by Gender (* is significant at p<.05)

Moderation by Gender	simple slop	t-value	Significance level (p)
Male	-1.391	-3.325	0.002*
Female	-0.927	-1.109	0.276*

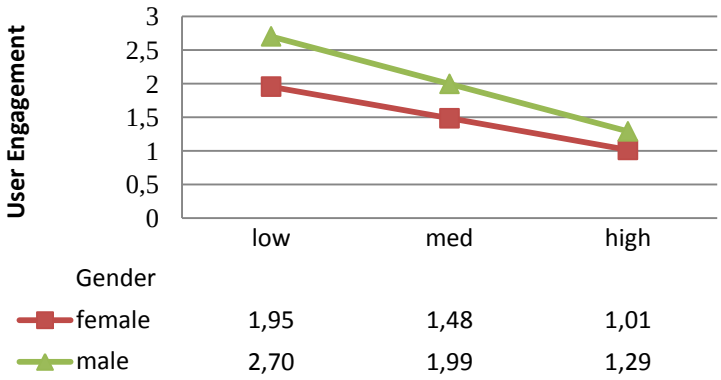


Figure 46 : Moderation by Gender

The male's slope was steeper than the female's. Figure 46 explains that the effect of the adaptive virtual character is more strongly associated with user engagement for males than for females. The adaptive virtual character is associated with higher user engagement under the condition that the user is a male.

Table 63 : Significance of slopes for moderation by Virtual Character Experiences (* is significant at p<.05)

Moderation by virtual character experiences	Simple slop	t-value	Significance level (p)
High	-1.369	-2.610	0.014*
Medium	-1.522	-1.170	0.252
Low	-1.675	-1.522	0.139

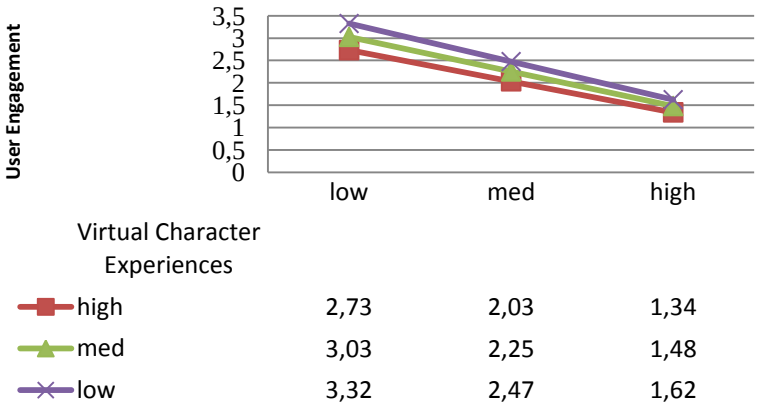


Figure 47 : Moderation by virtual character experiences

Figure 47 explains that previous experience with virtual characters is a buffer moderator, which refers to a decrease in the association between the conditions and user engagement. The effect of the adaptive virtual character is weakly associated with user engagement for users with high level of experience in virtual characters. The adaptive virtual character is associated with higher user engagement under the condition that the user had few experience before with virtual characters.

Table 64 : Significance of slopes for moderation by Ambient Experiences (* is significant at p<.05)

Moderation by Ambient Experiences	simple slop	t-value	Significance level (p)
High	-0.739	-2.227	0.034*
Medium	-0.344	-0.666	0.511
Low	0.052	0.069	0.945

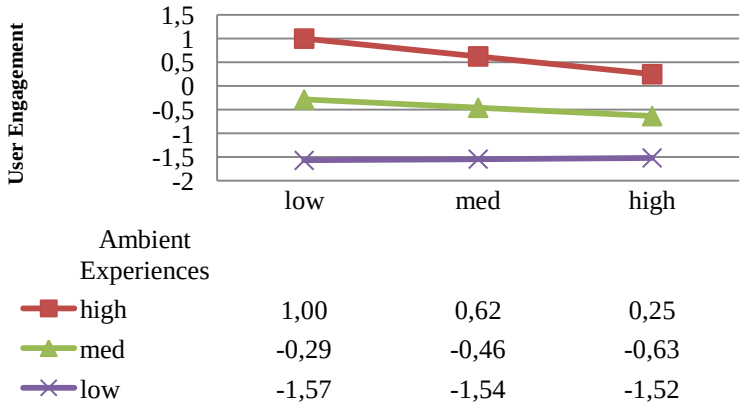


Figure 48 : Moderation by Ambient Experiences

Figure 48 explains that previous experience with ambient interaction is an amplifier moderator. It indicates an increase in the association between the conditions and user engagement. The effect of the adaptive virtual character is strongly associated with user's engagement when the user has a high level of experience with ambient environments. The adaptive virtual character is associated with higher user engagement under the condition that the user has a high level of experience with ambient environments.

In summary, gender and user expertise on virtual characters and ambient environments are both moderators. The associations between the conditions and user engagement are significantly different 1) between females and males, 2) among users with different levels of prior experiences with respect to virtual characters and ambient environments.

IV.2.10 Qualitative results

Users in the experimental condition reported few problems and felt proud of their experience with the adaptive virtual character. Users in the control condition provided more critical comments.

IV.2.10.1 Vocal help from the virtual character

The users in the control condition, who did not receive the visual cues of the adaptive virtual character (e.g. pointing or nodding) during the search tasks, reported a huge need to be able to speak to the virtual character. They argued that their visual attention was mostly focused on searching for objects and that they needed voice commands and speech synthesis. For example, "if it was possible to use voice commands, this could be very useful…it should be more auditory than visual…". This comment was not reported from users in the experimental condition. This is in line with the findings reported by Bellik et al (Bellik, Reba ïet al. 2009), according to which the output modality of a system can influence the input modalities that the user wants to use.

IV.2.10.2 Hesitation due privacy

We observed that users in the control condition hesitated a lot when opening the drawers to check if a target object was inside. We wondered why they were being so reluctant to do so. The reasons involve personality ("it is due to my personality, generally, I do not dare to do that, when I am at someone else's place"), affordability (I thought that the drawer could not be opened). In contrast, users in the experimental condition opened drawers more straightforwardly. The help provided by the adaptive virtual character might also be interpreted by users as a kind of social support, which breaks certain limitations with respect to privacy.

IV.2.10.3 Feelings and constructive suggestions

Users who interacted with the adaptive virtual character were more talkative and willing to give depth and close attention to the topic. The problems reported by users interacting with the adaptive virtual character, are feeling-related: "as the virtual character did not get what it wanted, it looked wicked", or "maybe you could include a more realistic environment around the virtual character to make the atmosphere less artificial". Users interacting with the adaptive virtual character were more involved in the discussion and provided constructive comments. For example: "it would be more interesting to have an immediate feedback, e.g. when we take the wrong direction, the virtual character should stop us right away", or "I found that the presence of the avatar in everyday life can bring much to those who live alone

and to seniors to help them find their objects easily". Users interacting with the adaptive virtual character felt that the experience was "interesting" and "fun", and largely accepted the experiment concept "it is the first-time experience like that and I really liked, personally I would like an avatar with me at home!"

IV.2.10.4 No perceived proximity

Though the users in the experimental condition reported satisfaction with respect to the adaptation of the virtual character in the questionnaire, no user confirmed that they did actually perceive the proximity changes of the virtual character within the screen.

Other studies showed the importance of proximity in human-human interaction and mediated interaction. Yet, our projection of the virtual character on a flat surface does not seem correctly to reflect this finding. This suggests that the distance perceived by the perspective effect might not be as important as the perceived distance from the physical interface. Finally, the virtual character's position in the physical settings in ambient environments might not be of importance in search tasks.

IV.2.11 Discussions

When designing a virtual character that adapts its spatial behavior to users' and objects' locations in a smart room, our assumption was that spatial nonverbal behaviors such as proximity, body orientation and posture might provide an intuitive aid to users in a spatial search task. Actually, the results support that the adaptive functions of the virtual character did increase the perceived presence and perceived adaptivity of the user, and did reduce the impact of task difficulty on their searching performance.

The framework focused on the effect of the adaptive virtual character on user perception (i.e. perceived adaptivity, perceived social presence, usability and satisfactions), the outcome of the task itself (i.e. performance) and the user nonverbal behavior (i.e. gazing, moving, and taking actions).

When investigating users' behavior in relation to their perception, we mapped the actual user actions to the perceptual criteria of evaluations. The correlations revealed that looking at the virtual character is likely to be related to perceive ease of use. Looking around is likely to be related to perceived control. The combinations of looking around during searching objects and looking at the virtual character during the instruction phase showed a high level of intrinsic motivation and positive affect towards the virtual character. Likeability is correlated to the duration of gaze towards specific objects in the ambient environment. This suggests that daring to take time to browse books or files during the search task benefits from a high level of likability towards the virtual character.

We investigated moderator and mediator effects based on the relationships between individual differences, the perceptions and the outcome of the interaction. We demonstrated how to analyze each type of effects using multiple regression and structural equation modeling technique. The results supported that the adaptive virtual character on the user engagement is mediated by user perceived social presence and moderated by gender and user prior expertise. The adaptive virtual character leads to feelings of social presence, which in turn, leads to high level of user engagement. The effect of the virtual character (adaptive or not) on user engagement depends on gender and prior experiences of the user.

IV.3 **Conclusions**

This chapter addressed the question about the selection of nonverbal spatial behaviors to be displayed by virtual characters in ambient environments and the relevant criteria for evaluation.

We designed and evaluated a location-aware virtual character that adapts its spatial behaviors to users' and objects' locations during a search task in a smart room.

We compared the behaviors and the reported perception of users interacting with this adaptive character to those of users interacting with a non adaptive character (i.e. that does not perceive nor use the location of users and objects).

We explained the process of evaluation with respect to user perception, user action and user characteristics. We combined several data and methods: files logged by the location detection system, indirect observations, questionnaires, and interviews.

First, we manually annotated user actions observed in the video. We then investigated the differences between the adaptive virtual character group and the control group at the behavioral level. Based on the correlations between different measured variables, we investigated the effects of the virtual character with respect to perceived social presence, user characteristics (gender and expertise) and engagement. The results showed that users interacting with the adaptive virtual character have a higher level of perceived adaptivity, perceived social presence and engagement than users interacting with the non-adaptive virtual character. Furthermore the interaction between the condition and the engagement is different between female and male users, but also between users with different levels of knowledge about ambient intelligence and virtual characters.

V. CONCLUSION AND FUTURE DIRECTIONS

V.1 Research contributions

Even if multimodal communication is key to smooth interactions between people, posture has been seldom explored compared to other modalities, such as speech and facial expressions. The postural expressions of others have a huge impact on how we situate and interpret an interaction both spatially and socially.

The preceding chapters have presented several contributions for laying the foundation for computational models that would relate postures to spatial and emotional communicative functions in human-computer interaction.

V.1.1 Interaction space in Real, Virtual and Real-Virtual world

The use of bodily interactions was assumed to enhance the conceptualization of space in users during interactions. Therefore, we studied bodily expressions within multiples interaction spaces and their potential use in human-computer interaction.

We investigated bodily expressions within space in the real world, space in a virtual environment and space in an ambient environment combining the virtual and real worlds.

Different areas were addressed, each having a specific view on how people employ space, bodily orientation and positioning as a means of organizing the attentional structure of social encounters. The results of several investigations showed that space in interaction could be studied via factors such as convergence, action tendencies and engagement in interaction.

Kendon (Kendon, Esposito et al. 2010) defined gesture space as a *transactional segment* taking the considerations of activities. A transactional segment is a space that extends in front of a person in which she carries out her activities. Hence, the context under which the activities occur is key to the size and the boundaries of this space. For example, during a face-to-face dyadic conversation, this space would be concentrated in front of the person. Telling a story using a monologue, this space might be larger (especially for the lower body). However, this transactional segment only involves the individual space.

Bodily Expressions in Real-world Adaptation and Human Interaction

Bodily adaptation during social interaction was also investigated. We defined postural convergence as the fact that the posture of the listener at time t adapts to the posture of the speaker at time t-1 within a shared space during dyadic conversation.

We developed a method to automatically analyze postural convergence in dyadic conversation based on manual annotations of postures. We observed that postural convergence mostly occurred in terms of arm swivel (maintain the arms close to the body or raise the arms), arm radial orientation (direct the arms aside or forward), leg distance (maintain the legs indented backward or straight) and leg swivel (maintain the legs outward bent at 90 °, or inward).

This study confirmed that posture may support the adaption between interlocutors in social interaction.

Bodily Expressions in Virtual Affect Display

The question about whether posture is a relevant modality to convey emotions involves the specification of reliable and discriminative features in bodily expressions of emotions.

We introduced an approach to explore how action tendencies can be expressed and perceived via bodily expressions. We simulated bodily expressions of action tendencies in a virtual character.

First, the forms of these emotional bodily expressions were investigated. We confirmed that bodily expressions might provide information about action tendencies.

Second, we investigated the extent to which the recognition of action tendencies varies as a function of the segmentation and duration of the animation of the virtual character. We found that the first two thirds of dynamic bodily expressions made already human viewers reached their perception threshold of action tendencies.

Three, we also observed other results that might be explored further in future studies:

1) the results suggested that human viewers would need more time to perceive the action tendency (Disappear from View) that relates to negative emotions (sadness, shame and guilt) than to perceive the action tendency (Attending) that related to neutral emotions (surprise and anxiety).

2) The results suggested there might be an intensity effect with respect to the segmentation and the duration of the animation and the perception of action tendencies. This means that the longer the human viewers were exposed to the dynamic bodily expressions of an action tendency, the stronger the emotions they attributed to that action tendency.

3) The results suggested that human viewers would be better at detecting threatening bodily expressions than at detecting non-threatening bodily expressions. This is in line with the observations from Neuroscience studies.

Bodily Expressions of Virtual Characters in Ambient Interaction with Human Users

We also investigated how users perceive bodily expressions of virtual characters in an ambient environment that combines the real and virtual worlds.

We designed a location-aware virtual character that uses bodily expressions to manage the shared space with users and real world objects.

The selection of bodily expressions accordingly was based on the location information and several task-related actions of users in the ambient environment. An experimental study was performed in an intelligent room involving 32 users. We investigated the effects of such a virtual character on users' behaviors (gaze, move actions and task-related taking actions) and user's perceptions (perceived ease of use, perceived efficiency, perceived control, emotional and attitudinal satisfaction, likeability, helpfulness, perceived social presence and perceived adaptivity).

The results showed that users interacting with such an adaptive virtual character had a higher level of perceived adaptivity, perceived social presence and engagement than users interacting with the non-adaptive virtual character.

V.1.2 Bodily expressions under *social* and *task-oriented* contexts

One of the important assumptions underlying our research is that context is an important component in human interaction as well as in human-computer interaction. The study that we described in this thesis involves social, task-oriented and mixed social-task contexts.

We explored the forms of bodily expressions by observing humans in several contexts, including spontaneous dyadic conversation, story-telling monologue and videos of users in the ambient environment. We collected several corpora and defined systematic coding schemes to code the postures (for both sitting and standing positions) but also to infer rules for their reproduction in virtual characters.

These studies under different contexts revealed different lines of research. First, in the case study of spontaneous dyadic conversation, we defined a systematic coding scheme to represent postures in such a social context. The posture was defined as any bodily changes within at least one second. This coding scheme describes the body postures with respect to four body parts. The annotations of postures were used to study how postures of the listener adapt to those of the speaker within a share space. This phenomenon of postural convergence was seen as a kind of bodily adaptation during social interaction.

Second, in the case study of story-telling monologues, we collected standing postures displayed during story-telling task. Our focus was to complete the previously defined coding scheme by including representations of leg postures when humans are in a standing position. We designed bodily expressions in virtual characters under a social context: a female virtual character expressed bodily expressions of action tendencies facing another static virtual character in a simulated social scene. We aimed to investigate whether users perceive the action tendencies and emotions in bodily expressions of virtual characters. We investigated the forms of these emotional bodily expressions, but also the extent to which the recognition of emotion varies as a function of the segmentation and duration of the animation. We reported the results of two perception tests addressing each of the issues.

Finally, we described how we designed the bodily expressions of a virtual character that interacted with users in ambient environments under a mixed task-oriented and social context. It involved a location-aware virtual character that uses bodily expressions to manage the shared space with users. The bodily expressions of this virtual character were made to help users in achieving a task. The selection of bodily expressions accordingly was based on the location information and several task-related actions of users in the ambient environment.

We distinguished the bodily expressions under social and task-oriented contexts, because their functions are meant to be different. For example, bodily expressions under social context can be in relation with social functions such empathy or emotions. In contrast, bodily expressions under task-oriented context basically refer to the performance of activities.

V.2 **Practical contributions**

There are several lessons learned from this work that are applicable to the practical matters of designing postural expressions in virtual characters and conducting human subjects experiments.

A user-centered experimental approach was applied throughout three phrases of my thesis work: observation of humans, perception by humans, and analysis of human behaviors in interaction. We were interested in three types of questions regarding the methodology: how to

validate the scheme that describes the bodily expressions, how to assess emotions of bodily expressions, and how to evaluate and model the spatial relationships that are based on bodily expressions between users and virtual characters.

V.2.1 Methodologies for Affective Assessment

We are interested in perceptual aspect of synthesized behavior. Studies on the analysis of affective bodily expressions mainly asked subjects to self-report if the bodily expressions match the verbal description of emotion components. Several studies also asked subjects to analyze the bodily expressions in terms of expressiveness qualities afterwards. However, the methodology concerning the perception of action tendencies is seldom discussed in the field.

Action tendency is the preparation phrase of changing relationships with the environments of humans, which are supposed to be subtle in bodily expressions. We asked whether human viewer report their perception of action tendencies through a self-reporting method using verbal description of action tendencies and emotional categories, as used in the study by Frijda (1989). Measures of action tendencies varied from adjective checklist to scales in our studies. Confidence level with respect to their judgment was jointly used. The goal of using these different methods was to judge which method would fit better the evaluation of action tendencies.

In the first experimental study (III.3), the two methods provided consist and complementary information about the perception of action tendencies based on the static postures. In the second experimental study (III.5), the two methods reported different resultants with respect to the perception of action tendencies based on the dynamic bodily expressions. This suggests that methods of evaluation of action tendencies should be chose in reference to the characteristic of bodily expressions.

V.2.2 Methodologies for Ambient Interaction Design

Criteria of evaluations

We selected several criteria for evaluations of our humanoid ambient display in terms of perceptual and behavioral evaluations. Perceptual evaluations showed that a virtual character that adapted its bodily expressions to the locations of users and objects made users feel like being the presence of another human being. Behavioral evaluations showed that such a virtual character made users engaged in the interaction based on their gaze behaviors. Perceived social presence and engagement were seen as relevant criteria to evaluate the ambient interaction design that involves virtual characters.

Third-variable modeling

Designing interaction in ambient environments requires considering human factors such as user characteristics, user perceptions and user behaviors. We considered third variables (moderator and mediator) in modeling the relationships between the virtual character and its impact on users. Moreover, there are two factors that moderate the association of virtual characters and user behavior in terms of engagement: gender and user expertise with respect to virtual characters and ambient environments. The associations are significantly different 1) between females and males, and 2) among users with different levels of prior experiences with respect to virtual characters and ambient environments. First, the adaptive virtual character was associated with higher user engagement under conditions that users are males. Second, the adaptive virtual character was associated with higher user engagement under conditions that users lack experiences with respect to virtual characters. Third, the adaptive

virtual character was associated with higher user engagement under conditions that user had rich experiences with respect to ambient environments.

V.3 **Future directions**

Motion capture technique

The manual annotation of videos for designing nonverbal behavior in virtual characters is commonly questioned about the precision of the data. Motion capture data can provide precise information about body movements (Kleinsmith and Bianchi-Berthouze 2007). Motion capture uses machine learning methods to infer the role of posture in emotion expressions. Motion capture has advantages such as efficacy with respect to data collection, accuracy of the collected data and flexibility to transform the collected data into virtual characters. For example, motion capture would enable capturing the subtle body movement in the preparation of emotional actions (action tendency).

Using motion caption systems raises practical questions about selecting relevant methods of machine learning to map discriminative posture features to the emotions (ex. categories of action tendencies). Several analysis methods can be used to address this question: Mixtures of Gaussian (use several Gaussian distributions to model different groups), Neural Network (use a three-layer perception network, where each layer has its own weights and bias functions, to classify data into six different postures) and Support Vector Machines (use a nonlinear mapping kernal to project data onto a higher dimension in order to find a separating hyper plane with the largest margin between the data). Other methods also estimate parameters and select models: ML (Maximum-Likelihood), EM (Expectation-Maximization), Entropy (Maximum Entropy Discrimination) and MAP (Maximum-a-posteriori). These methods enable to measure the saliency of the features in perception of emotions by discriminating between two or more categories or groups.

Individual differences

Another future direction involves the investigation of posture with respect to individual expressive profiles. We found several elements that suggested individual differences in relation with postural expressions. For examples, the clustering analysis presented in II.2.3 and II.4.5 suggested that only a small set of postures are commonly used by each of the subjects during dyadic conversation and monotony story-telling. We observed an obvious difference between two interlocutors during dyadic conversation with respect to postural convergence in II.3. Moreover, the experimental study that is reported in IV.2 showed that the effect of the adaptive virtual character is more sternly associated with user engagement for males than for females (IV.2.9.3.2.2).

This direction is expected to enable to investigate other individual differences such as personality with respect to expression and perception of bodily signs. Cross-validation techniques are used in the literature to estimate how well the model learned from the encoder perception is mapped to the model learned from the decoder. For examples, the common cross-validation techniques include repeated random sub-sampling validation, K-fold cross-validation, k $\times$ 2 cross-validations and leave-one-out cross-validation.

Studying social interaction in real, virtual and real-virtual worlds

People perform different postural expressions according to how they arrange themselves spatially in various kinds of contexts where people search to meet their interactional needs (Kendon, Esposito et al. 2010).

The results of the convergence study can be applied to modeling postural expressions of multiple virtual characters during the display of virtual social scenes. The bodily expressions can be incorporated with other multimodal signals, such as facial expressions to model social relationships between virtual characters (Pedica, Hogni et al. 2009).

The contributions of this thesis can also be useful for multi-user and multi-agents interactions in complex ambient environments mixing various notions of space (Figure 49).

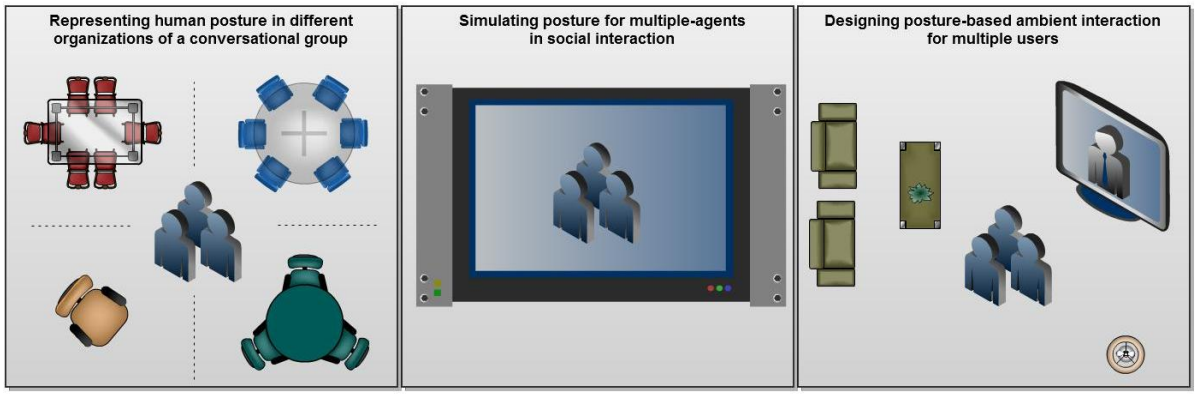
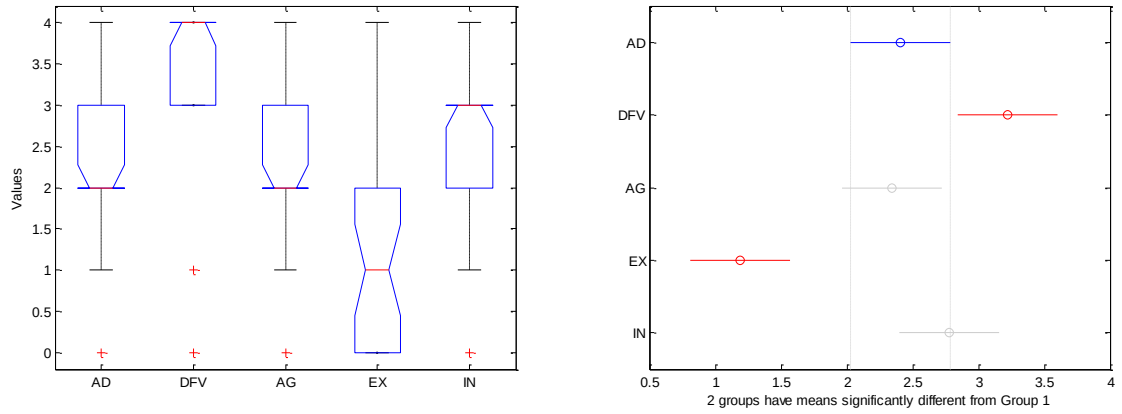


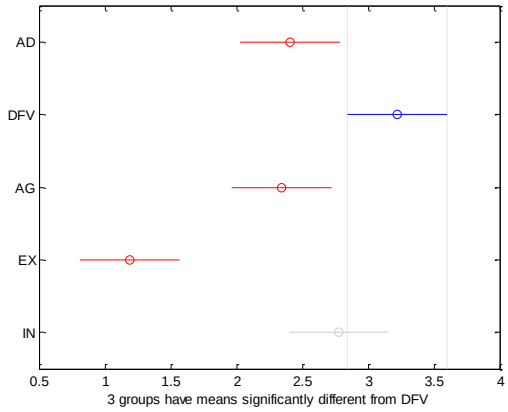
Figure 49: Future directions of the work

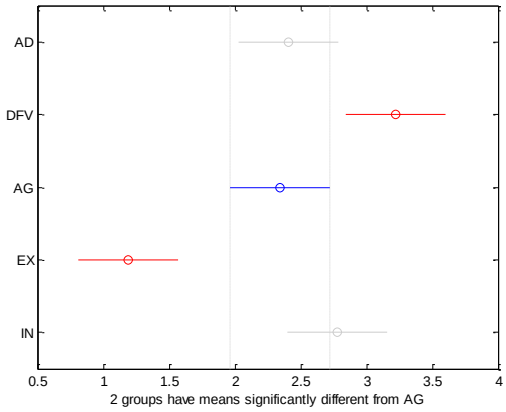
APPENDICES

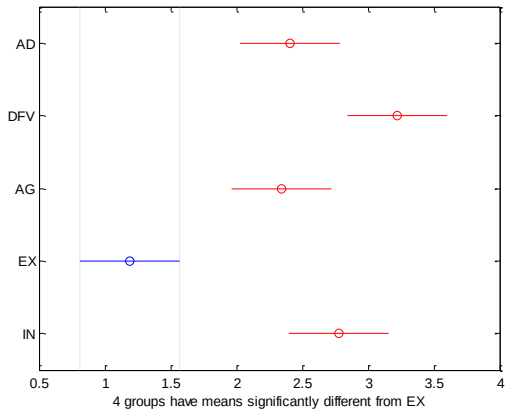
Appendix 1. Statistical results in the experimental study on dynamic bodily expressions of action tendencies.

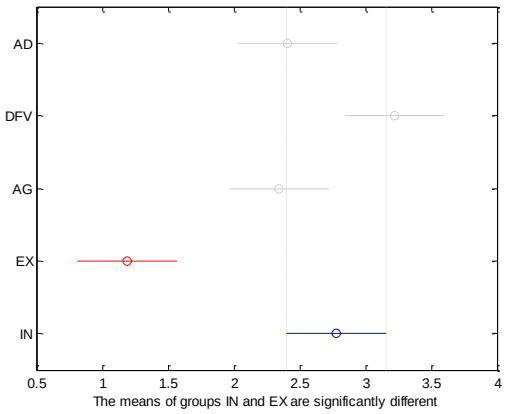
Within the condition 1, evaluation scores of action tendencies were significantly different each other, F (4, 155) = 14.88, MS = 18.306, p<.05. The mean score of Disappear from View (M= 3.2188, SB=0.1961) was significantly higher than Attending (M= 2.4063, SB= 0.1961), Antagonistic (M= 2.3438, SB=0.1961) and Exuberant (M= 1.1875, SB=0.1961). The mean score of Attending (M= 2.4063, SB= 0.1961), Disappear from View (M= 3.2188, SB=0.1961), Antagonistic (M= 2.3438, SB=0.1961) and In command (M= 2.7813, SB=0.1961) was respectively significantly higher than that of Exuberant (M= 1.1875, SB=0.1961).



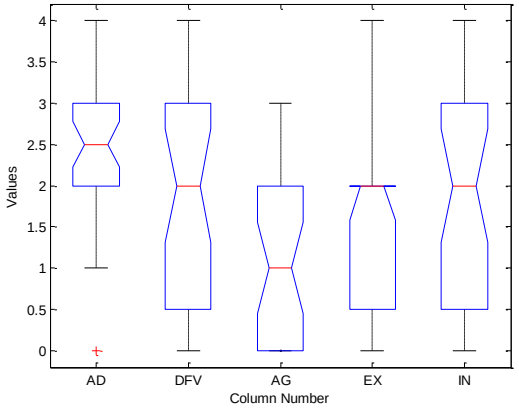


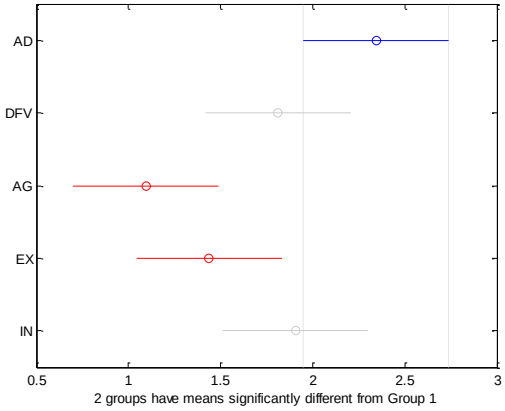


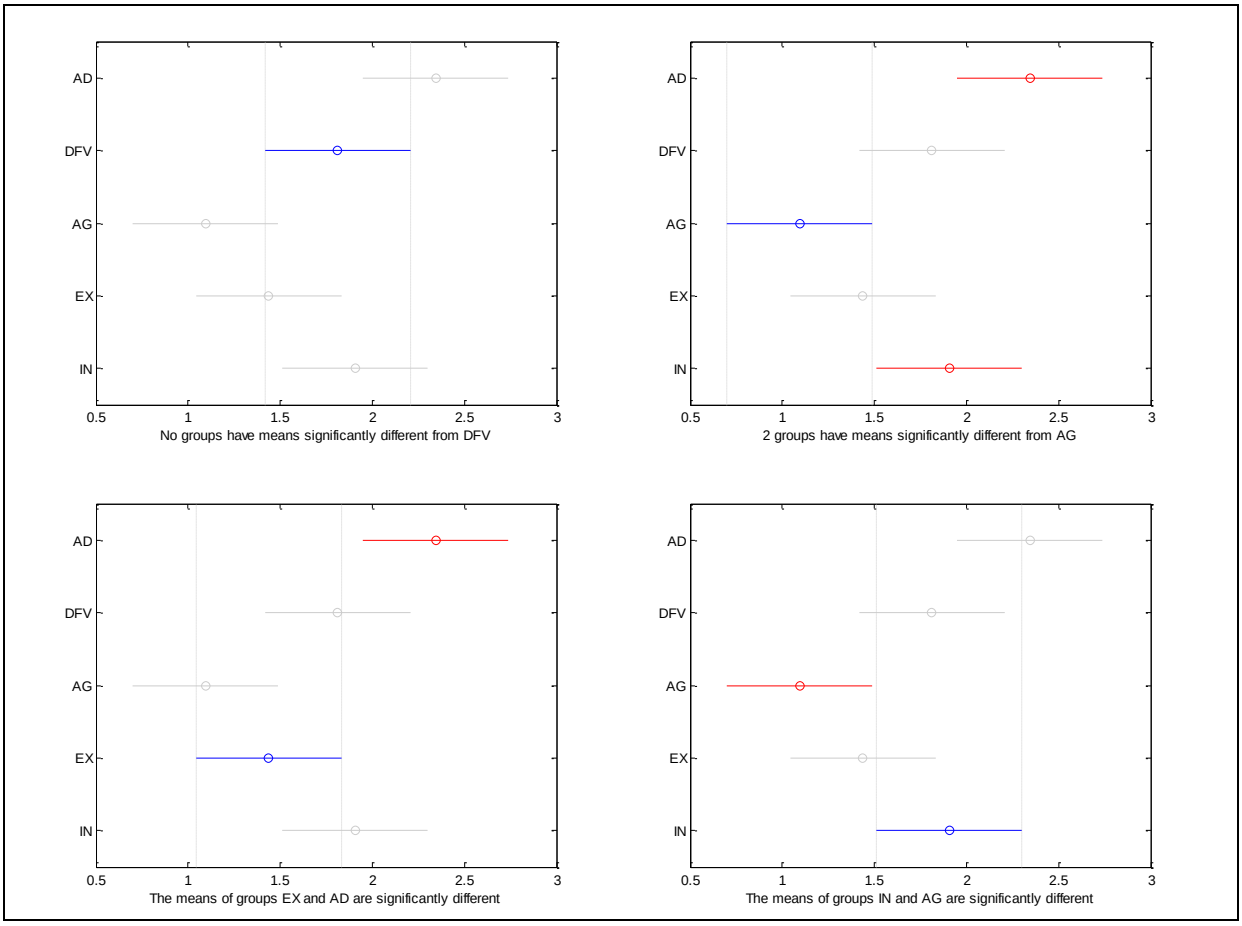




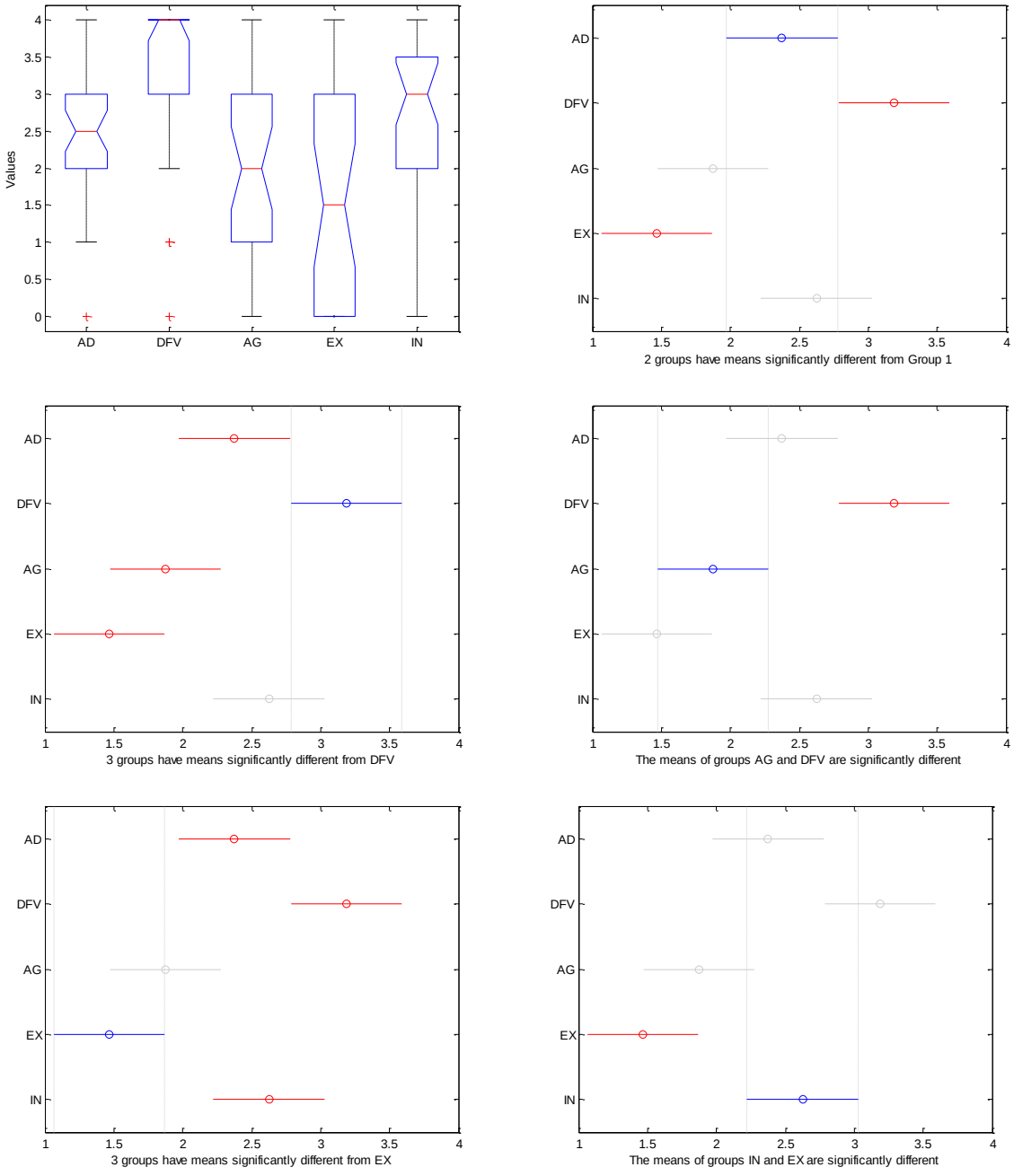
Within the condition 2, F (4, 155) = 5.41, MS= 7.234, p<.05. Post hoc analyses showed that the mean score of perceived Attending (M=2.3438, SD=0.2045) was significantly higher than Antagonistic (M=1.093, SD=0.2045) and Exuberant (M=1.438, SD=0.205). The mean score of In command (M=1.9063, SD=0.205) was significantly higher than that of Antagonistic (M=1.093, SD=0.2045).






AD
DFV
AG
EX
IN
0.5
1
1.5
2
2.5
3
No groups have means significantly different from DFV
2 groups have means significantly different from AG
The means of groups EX and AD are significantly different
The means of groups IN and AG are significantly different

Within the condition 3, F (4, 155) =, MS = 14.163, p<.05. The mean score of Disappear from View (M= 3.188, SD=0.208) was significantly higher than that of Attending (M= 2.375, SD= 0.208), Antagonistic (M=1.8750, SD=0.208) and Exuberant (M=1.4688, SD=0.208). The mean score of In command (M=2.6250, SD=0.208) was significantly lower than that of Exuberant (M=1.4688, SD=0.208).



The interaction of Segmentation and Duration of the Animation and action tendencies was significant, F (8, 465) = 3.76, MS = 4.963, p < .05, indicating that the effect of Segmentation and Duration of the Animation was different from some action tendencies to others. We

performed univariate analyses within each of action tendencies, with Segmentation and Duration of the Animation as within-subjects factor and evaluation scores of action tendencies as the dependent variable.

For the action tendency Attending, no significant difference was found among levels of Segmentation and Duration of the Animation, F (2, 93) =0.03, MS=0.03, p>.05. The scores of perceived action tendencies were relatively high across levels of Segmentation and Duration of the Animation for both condition 1 (M=2.406, SB =0.176), condition 2 (M=2.344 SB =0.176) and condition 3 (M=2.375, SB =0.176).

For the action tendency Disappear from View, F (2, 93) =13.56, MS= 20.64, p<.05, the mean scores for condition 2 (M=1.813, SB = 0.218) was significantly lower than for condition 1 (M=3.219, SB =0.218) and condition 3 (M=3.188, SB =0.218). No significant difference was found between condition 1 and condition 3.

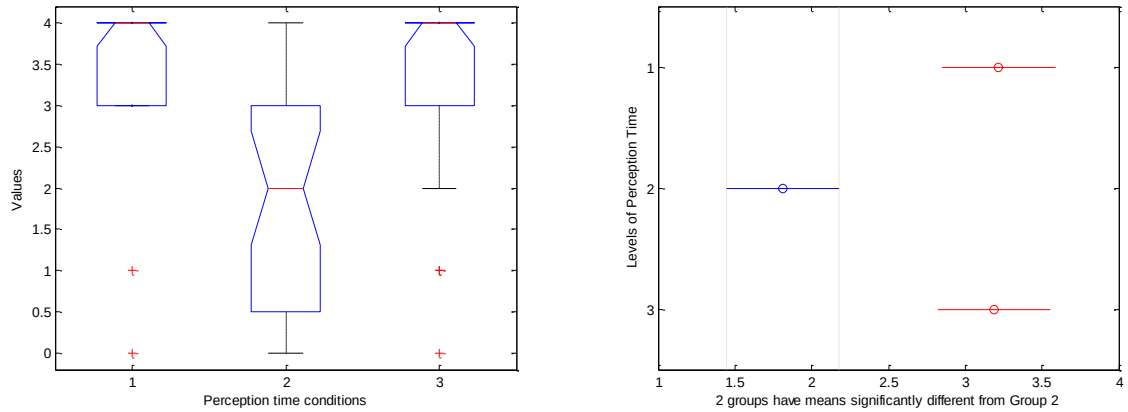


Figure 50: Effect of Segmentation and Duration of the Animation within the action tendency Disappear from View **(comparison of mean scores of perceived action tendencies).**

For the action tendency Antagonistic, F (2, 93) =10.65, MS= 12.76, p<.05, the mean scores for condition 2 (M=1.094, SB = 0.194) was significantly lower than for condition 1 (M=2.344, SB =0.194) and condition 3 (M=1.875, SB =0.194).

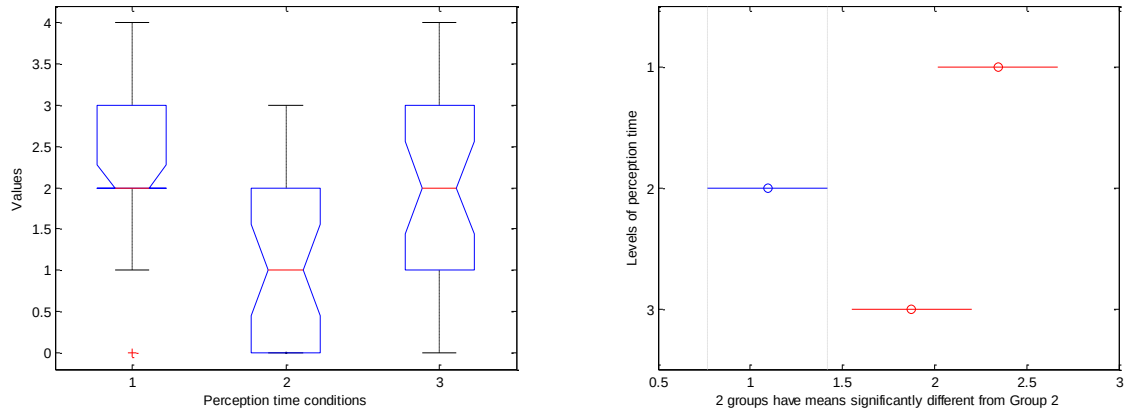


Figure 51: Effect of Segmentation and Duration of the Animation within the action tendency Antagonistic **(comparison of mean scores of perceived action tendencies).**

For the action tendency Exuberant, no significant difference was found among levels of Segmentation and Duration of the Animation, F (2, 93) =0.52, MS=0.760, p>.05. The scores of perceived action tendencies were generally low across levels of Segmentation and Duration of the Animation for both condition 1 (M=1.188, SB =0.213) , condition 2 (M=1.438, SB =0.213) and condition 3 (M=1.469, SB =0.213).

For the action tendency in command, F (2, 93) =4.85, MS= 6.969, p<.05, the mean scores for condition 2 (M=1.906, SB =0.2119) was significantly lower than that for condition 1 (M=2.781, SB =0.2119) and condition 3 (M=2.625, SB =0.2119).

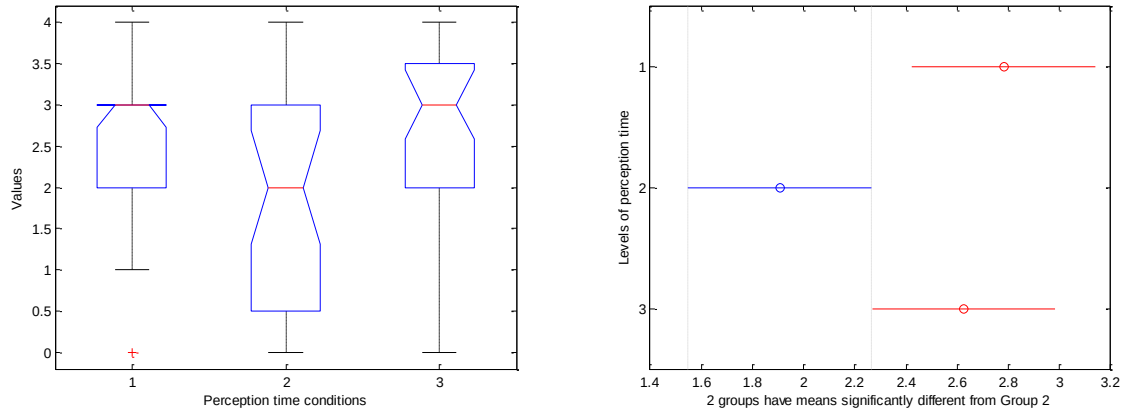


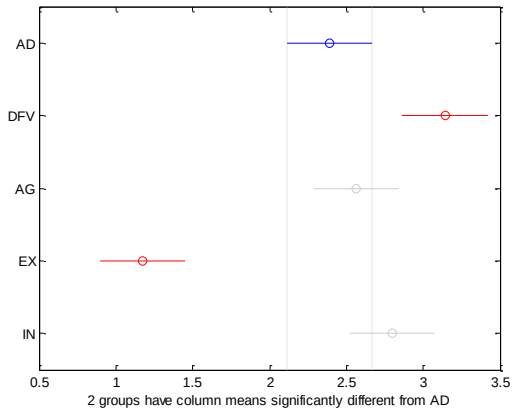
Figure 52: Effect of Segmentation and Duration of the Animation within the action tendency in command (comparison of mean scores of perceived action tendencies).

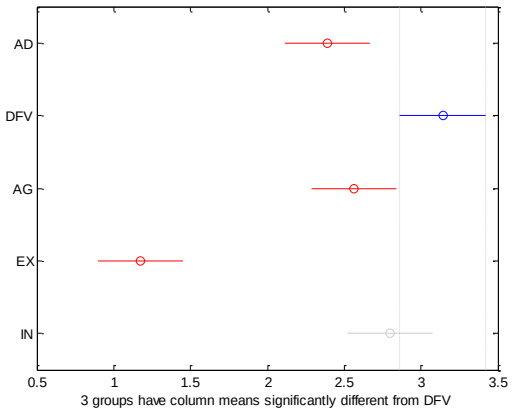
Appendix 2. Detailed results of the evaluations of the dynamic postural expressions of action tendencies

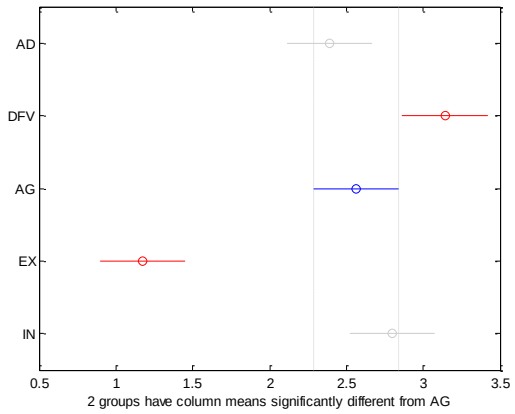
A. Effect of Movement Quality on Perceived Action Tendencies

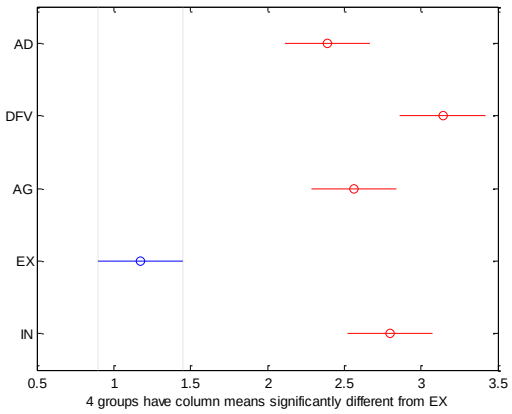
We performed a one-way ANOVA, with Movement Quality as within-subjects factors and perceived action tendencies as the dependent variable. The main effect of Movement Quality had no significance on perceived action tendencies, $F(1,310) = 0.698$, $MS = 0.2$, $p > .05$.

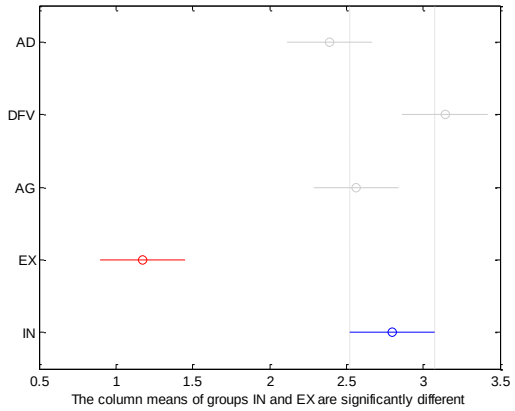
Evaluation scores of action tendencies were significantly different each other across overall levels of Movement Quality, $F(4, 310) = 27.05$, $p<.05$. Post hoc analyses showed that the mean score of Disappear from View (M= 3.1406, SD= 0.1439) was significantly higher than that of Attending (M= 2.3906, SD= 0.1439), Antagonistic (M= 2.5625, SD= 0.1439), and Exuberant. (M= 1.1719, SD= 0.1439). The mean score of Attending (M= 2.3906, SD= 0.1439), Antagonistic (M= 2.5625, SD= 0.1439) and In command (M= 2.7969, SD= 0.1439) was significantly higher than that of Exuberant (M= 1.1719, SD= 0.1439).







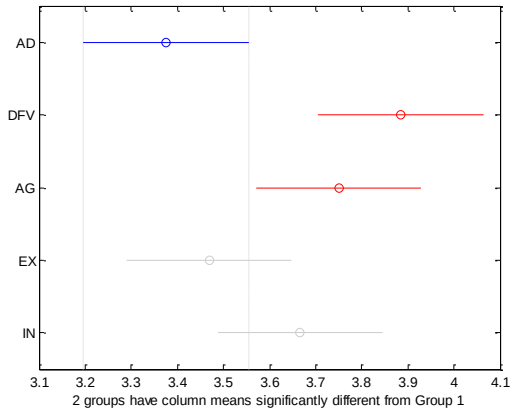


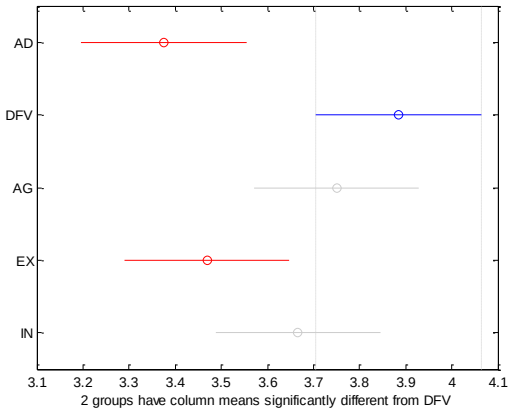


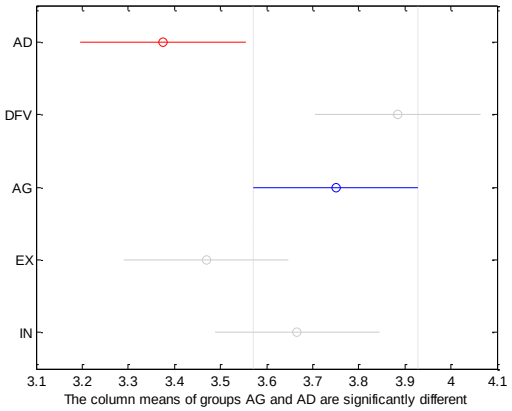
B. Effect of Segmentation and Duration of the Animation on Confidence

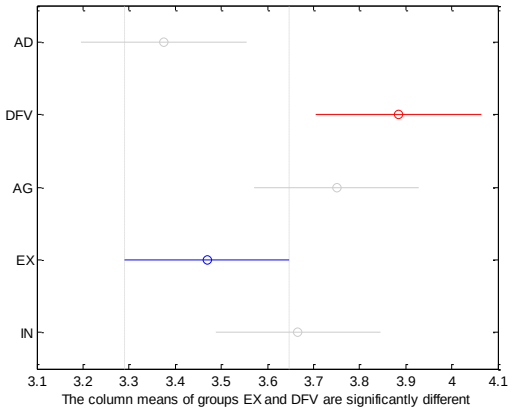
The main effect of Segmentation and Duration of the Animation was significant on confidence of judgments, F (2, 465) =21.26, MS=17.665, p<.05. Post hoc analyses showed that the mean score in condition 2 (M=3.2563, SD=0.0721) was significantly lower than in condition 1 (M=3.8937, SD=0.0721) and in condition 3 (M=3.7375, SD=0.0721).

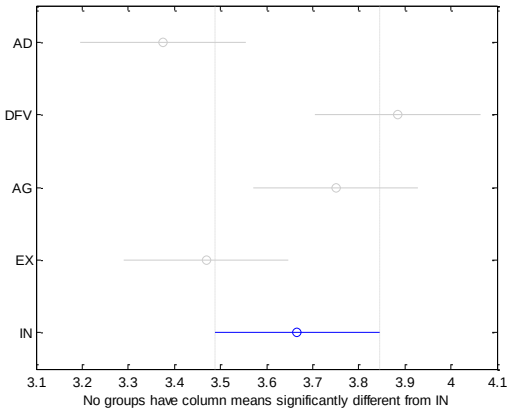
Confidence scores of action tendencies were significantly different each other across overall levels of Segmentation and Duration of the Animation, F (4, 465) = 4.97, MS=4.128, p<.05. Post hoc analyses showed that the mean score of Attending (M=3.3750, SD=0.0930) was significantly lower than that of Disappear from View (M=3.8854, SD=0.0930) and Antagonistic (M=3.7500, SD=0.0930). The mean score of Exuberant (M=3.4688, SD=0.0930) was significantly lower than that of Disappear from View (M=3.8854, SD=0.0930).






AD
DFV
AG
EX
IN
3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4 4.1
The column means of groups AG and AD are significantly different


AD
DFV
AG
EX
IN
3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4 4.1
The column means of groups EX and DFV are significantly different


AD
DFV
AG
EX
IN
3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4 4.1
No groups have column means significantly different from IN

Appendix 3. Dummied values for data analysis of posture annotations using EXPO-sitting scheme

Postures.Arms.LeftArm:ArmsHeight

****

0 = none

1 = Above head

2 = Head

3 = Shoulder

4 = Chest

5 = Abdomen

6 = Waist

7 = Hip/buttock

8 = Thigh

Postures.Arms.LeftArm:ArmsDistance

****

0 = none

1 = Far

2 = Normal

3 = Close

4 = Touch

5 = Backup

Postures.Arms.LeftArm:ArmsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

6 = Inside

Postures.Arms.LeftArm:ArmsRadialZ

****

0 = none

1 = Forward

2 = Obverse

3 = Downward

4 = Reverse

5 = Backward

6 = Upward

Postures.Arms.LeftArm:ArmsSwivel

****

0 = none

1 = Touch

2 = Normal

3 = Out

4 = Orthogonal

5 = Raised

Postures.Arms.LeftArm:ForearmHandOrientation

****

0 = none

1 = Palm up

2 = Palm down

3 = Palm towards self

4 = Palm away from self

5 = Palm towards addressee

6 = Palm on side inwards

7 = Palm on side outwards

Postures.Arms.LeftArm:ArmsTouch

****

0 = none

1 = Head

2 = Arm

3 = Trunk

4 = Leg

5 = Furniture

6 = Clothes

7 = Nottouching

Postures.Arms.RightArm:ArmsHeight

****

0 = none

1 = Above head

2 = Head

3 = Shoulder

4 = Chest

5 = Abdomen

6 = Waist

7 = Hip/buttock

8 = Thigh

Postures.Arms.RightArm:ArmsDistance

****

0 = none

1 = Far

2 = Normal

3 = Close

4 = Touch

5 = Backup

Postures.Arms.RightArm:ArmsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

6 = Inside

Postures.Arms.RightArm:ArmsRadialZ

****

0 = none

1 = Forward

2 = Obverse

3 = Downward

4 = Reverse

5 = Backward

6 = Upward

Postures.Arms.RightArm:ArmsSwivel

****

0 = none

1 = Touch

2 = Normal

3 = Out

4 = Orthogonal

5 = Raised

Postures.Arms.RightArm:ForearmHandOrientation

****

0 = none

1 = Palm up

2 = Palm down

3 = Palm towards self

4 = Palm away from self

5 = Palm towards addressee

6 = Palm on side inwards

7 = Palm on side outwards

Postures.Arms.RightArm:ArmsTouch

****

0 = none

1 = Head

2 = Arm

3 = Trunk

4 = Leg

5 = Furniture

6 = Clothes

7 = Nottouching

Postures.Shoulder:ShoulderType

****

0 = none

1 = Raise left shoulder

2 = Raise right shoulder

3 = Raise shoulders

4 = Lower left shoulder

5 = Lower right shoulder

6 = Lower shoulders

7 = Move shoulder forward

8 = Move shoulder back

9 = Shoulder shrug

10 = Shoulder twist

Postures.Trunk:type

****

0 = none

1 = Lean forward

2 = Lean back

3 = Turn toward person

4 = Turn away from person

5 = Lean toward person

6 = Lean away from person

7 = Lower trunk

8 = Raise trunk

Postures.Legs.LeftLeg:LegsHeight

****

0 = none

1 = Chest

2 = Abdomen

3 = Belt

4 = Buttock

5 = Thigh

Postures.Legs.LeftLeg:LegsDistance

****

0 = none

1 = Feet behind Knee

2 = Feet in front of Knee

Postures.Legs.LeftLeg:LegsSwivel

****

0 = none

1 = Feet outside Knee

2 = Feet inside Knee

Postures.Legs.LeftLeg:LegsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

Postures.Legs.LeftLeg:LegToLegDistance

****

0 = none

1 = Knees apart Ankles together

2 = Knees together Ankles apart

3 = Knees together Ankles together

4 = Knees apart Ankles apart

Postures.Legs.LeftLeg:CrossedLegs

****

0 = none

1 = Ankle over thigh

2 = At knees

3 = At ankles

4 = Feet over feet

5 = Cross legged

Postures.Legs.RightLeg:LegsHeight

****

0 = none

1 = Chest

2 = Abdomen

3 = Belt

4 = Buttock

5 = Thigh

Postures.Legs.RightLeg:LegsDistance

****

0 = none

1 = Feet behind Knee

2 = Feet in front of Knee

Postures.Legs.RightLeg:LegsSwivel

****

0 = none

1 = Feet outside Knee

2 = Feet inside Knee

Postures.Legs.RightLeg:LegsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

Postures.Legs.RightLeg:LegToLegDistance

****

0 = none

1 = Knees apart Ankles together

2 = Knees together Ankles apart

3 = Knees together Ankles together

4 = Knees apart Ankles apart

Postures.Legs.RightLeg:CrossedLegs

****

0 = none

1 = Ankle over thigh

2 = At knees

3 = At ankles

4 = Feet over feet

5 = Cross legged

Appendix 4. Dummied values for analyzing postural convergence using the EXPO-standing scheme

turntype

****

0 = none

1 = speaker

2 = listener

ArmsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

6 = Inside

ForearmHandOrientation

****

0 = none

1 = Palm up

2 = Palm down

3 = Palm towards self

4 = Palm away from self

5 = Palm towards addressee

6 = Palm on side inwards

7 = Palm on side outwards

ArmsTouch

****

0 = none

1 = Head

2 = Arm

3 = Trunk

4 = Leg

5 = Furniture

6 = Clothes

7 = Nottouching

ArmsSwivel

****

0 = none

1 = Touch

2 = Normal

3 = Out

4 = Orthogonal

5 = Raised

ArmsHeight

****

0 = none

1 = Above head

2 = Head

3 = Shoulder

4 = Chest

5 = Abdomen

6 = Waist

7 = Hip/buttock

8 = Thigh

ArmsRadialZ

****

0 = none

1 = Forward

2 = Obverse

3 = Downward

4 = Reverse

5 = Backward

6 = Upward

ArmsDistance

****

0 = none

1 = Far

2 = Normal

3 = Close

4 = Touch

5 = Backup

ArmsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

6 = Inside

ForearmHandOrientation

****

0 = none

1 = Palm up

2 = Palm down

3 = Palm towards self

4 = Palm away from self

5 = Palm towards addressee

6 = Palm on side inwards

7 = Palm on side outwards

ArmsTouch

****

0 = none

1 = Head

2 = Arm

3 = Trunk

4 = Leg

5 = Furniture

6 = Clothes

7 = Nottouching

ArmsSwivel

****

0 = none

1 = Touch

2 = Normal

3 = Out

4 = Orthogonal

5 = Raised

ArmsHeight

****

0 = none

1 = Above head

2 = Head

3 = Shoulder

4 = Chest

5 = Abdomen

6 = Waist

7 = Hip/buttock

8 = Thigh

ArmsRadialZ

****

0 = none

1 = Forward

2 = Obverse

3 = Downward

4 = Reverse

5 = Backward

6 = Upward

ArmsDistance

****

0 = none

1 = Far

2 = Normal

3 = Close

4 = Touch

5 = Backup

ShoulderType

****

0 = none

1 = default

2 = Raise left shoulder

3 = Raise right shoulder

4 = Raise shoulders

5 = Lower left shoulder

6 = Lower right shoulder

7 = Lower shoulders

8 = Move shoulder forward

9 = Move shoulder back

type

****

0 = none

1 = default

2 = Lean forward

3 = Lean back

4 = Turn toward person

5 = Turn away from person

6 = Lean toward person

7 = Lean away from person

8 = Lower trunk

9 = Raise trunk

LegToLegDistance

****

0 = none

1 = Knees apart Ankles together

2 = Knees together Ankles apart

3 = Knees together Ankles together

4 = Knees apart Ankles apart

LegsHeight

****

0 = none

1 = Chest

2 = Abdomen

3 = Belt

4 = Buttock

5 = Thigh

CrossedLegs

****

0 = none

1 = Ankle over thigh

2 = At knees

3 = At ankles

4 = Feet over feet

5 = Cross legged

6 = non crossement

LegsSwivel

****

0 = none

1 = Feet outside Knee

2 = Feet inside Knee

3 = Feet in line with Knee (default)

LegsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

LegsDistance

****

0 = none

1 = Feet behind Knee

2 = Feet in front of Knee

3 = Feet in line with Knee

LegToLegDistance

****

0 = none

1 = Knees apart Ankles together

2 = Knees together Ankles apart

3 = Knees together Ankles together

4 = Knees apart Ankles apart

LegsHeight

****

0 = none

1 = Chest

2 = Abdomen

3 = Belt

4 = Buttock

5 = Thigh

CrossedLegs

****

0 = none

1 = Ankle over thigh

2 = At knees

3 = At ankles

4 = Feet over feet

5 = Cross legged

6 = non crossement

LegsSwivel

****

0 = none

1 = Feet outside Knee

2 = Feet inside Knee

3 = Feet in line with Knee (default)

LegsRadialOrientation

****

0 = none

1 = Behind

2 = Out

3 = Side

4 = Front

5 = Inward

LegsDistance

****

0 = none

1 = Feet behind Knee

2 = Feet in front of Knee

3 = Feet in line with Knee

Appendix 5. Collecting a scenario-based corpus for designing bodily expressions of virtual characters

In this section of appendix, we present an illustrative scenario-based design process involving video corpora. First, we collected a corpus of bodily expressions according to a scenario occurring in ambient environments. Then, we designed the bodily expressions of an adaptive virtual character based on this video corpus.

A. Context

The test bed for research for the ATRACO project is the Essex intelligent apartment (iSpace) which is shown in Figure 53. The iSpace is a spacious two bedroom flat with a kitchen and a bathroom, which provides the possibility for examining the deployment of embedded agents and ambient user interfaces. A scenario of "Prepare dinner for friends" was applied into iSpace, as shown in Figure 53.



Figure 53: Virtual characters are used as ambient interaction metaphor in iSpace[4]

B. Objective

The objective was to design a virtual character that adapts its bodily expressions depended on the adaptation criteria that are relevant to ambient interaction. We revealed three adaptation criteria that were abstracted from the user interaction context. We then selected the bodily cues based on the following static and dynamic factors.

- **Personality** preferred by the user (static factor): the virtual character should adapt its nonverbal cues accordingly to users' preferences regarding extraversion / introversion.
- **Communicative acts** (dynamic factor) engaged by the virtual character: we consider communicative acts as dynamically changing parameters of the user interaction with the ambient system, especially of the human-computer conversation. For example, the agent makes different gestures according to communicative function - inform, confirm and warning-. (Cassell, Torres et al. 1999) indicated that nonverbal conversational signals of the virtual character such as averting gaze and lifting eyebrows when taking turn, or performing beat

[4] http://iieg.essex.ac.uk

gestures when providing content, are of great importance for the user. What we should know about the taxonomy of DIT++ with respect to DAMSL (Core and Allen 1997) is that a number of important communicative functions such as questions, answers, request, and statements are set apart of dimensions-specific functions as general-purpose functions in the sense of that they can be used in any dimension, depending on their semantic content.

- **The audio availability** of the user (dynamic factor): if the user is not available in his audio channels, the virtual character would rather use nonverbal behavior than speech.

C. Scenario: "Prepare dinner for friends"

The core of scenario is a narrative around a user, trying to achieve a task goal within a given context or environment. We describe briefly in this section a scenario of preparing dinner for friends.

> The user starts cooking the pasta. He fills the cooking pot with water and places it on the kitchen stove, which automatically switches on to the correct temperature, according to the recipe. The user occupies himself with other cooking tasks in the kitchen. When the water in the cooking pot has almost reached the boiling temperature, the virtual character locates the user (he is nearby, in the kitchen), informs him to add some salt and warns him that in a couple of minutes, he'll have to add the pasta. A couple of minutes later, the user adds the pasta; the virtual character detects that the user has lifted the pasta from the kitchen top and infers that he is going to add them. The virtual character confirms with the user that pasta has been added and informs him that, given the current stove temperature, the pasta will boil within twelve minutes.
>
> The user decides to place a phone call to his parents. He moves to the living room corner and asks the virtual character not to interrupt him. After a while, an alarm that the pasta is almost cooked is produced. The virtual character decides to inform the user but also wants to respect the user's non-interruption request. The virtual character uses alternative communication means, while at the same time, instructs the stove to lower the cooking temperature.

We planned the followings steps based on the described scenario (Figure 54) to conduct the collection of a corpus of bodily expressions.

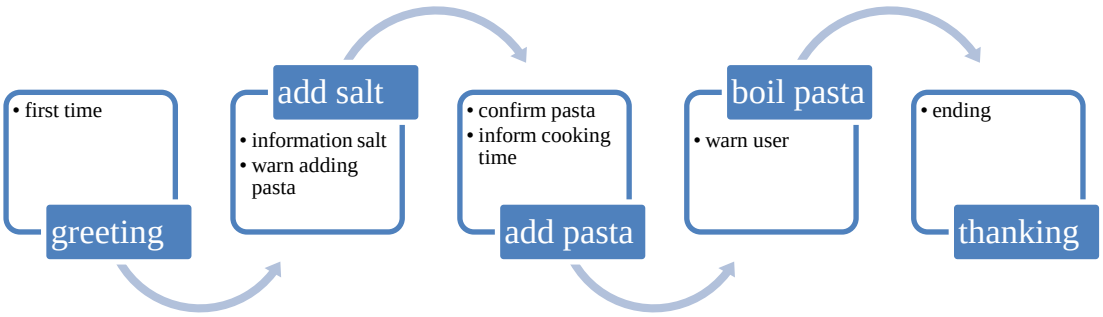


Figure 54: Steps for the role-playing

D. Episodes

We selected several episodes (as listed in order below) based upon which the actor was asked to play the role of the adaptive virtual character.

The episodes are related to a set of communicative acts to each of them a concise definition is given. Bearing this basic knowledge and the scenario descriptions in mind, the actor had to speak (verbal) and manifest the body language (nonverbal) accordingly to express himself within the different episodes of the scenario.

In the episode "warning the user that the cooking time almost over", the actor should play the role by two different means of communication: 1) only use body language 2) use both of words and nonverbal cues, such as postures, gestures, body orientation and distance.

Communicative acts	Episodes in the scenario
Greeting first time: in the first time the virtual character interacting with the user will give a self-introduction and then introduce the system.	Welcome
*Greeting later: the virtual character will greet the user if he hasn't seen him for more than 3 hours or just ask if everything is ok.	Welcome
Inform: The speaker believes that the information he provides is correct.	The virtual character informs the user to add some salt.
Warning: The speaker evaluates the situation, described in the semantic content, to be dangerous or potentially harmful for the addressee.	The virtual character warns that in a couple of minutes, the user will have to add the pasta.
Confirm: The speaker believes that Addressee believes that the propositional content is true.	The virtual character confirms that the pasta has been added.
Inform: The speaker believes that the information he provides is correct.	The virtual character informs that given the current stove temperature, the pasta will be cooked within twelve minutes.
Warning: The speaker evaluates the situation, described in the semantic content, to be dangerous or potentially harmful for Addressee.	The virtual character warns that the cooking time is almost over and the stove will automatically switch off.
*Gratitude in between: at the end of the subtask, the virtual character will show that he is grateful for having done the job together with the user.	
Gratitude end: at the end of using experience, the virtual character would like to show that he is grateful for having shared experience with the user.	

Table 65: The main episodes and the related communicative acts

The scenario "prepare dinner for friends" [5] mainly involves two parts of interaction tasks between the system and the user: negotiation and collaboration. Only the collaboration part will be instantiated with the virtual character. Not only the technical considerations, but also for the reason that virtual character as a multimodal user interface using natural human communication cues in output is expected to stimulate face-to-face communication during the human-computer interaction which leads to higher levels of cooperation.

E. Specifications

- Personality

Personality	Nonverbal behaviors	Distinctive parameters
Extrovert	make wide movements, use expansive gestures, pose limbs spread wide from its body, approach more freely others in space	Spatial extension Gesture units Velocity of movements Fluidity Handedness
Introvert	keep limbs close to the body, gesture less freely, avoid approaching others in space	

Table 66: Distinctive nonverbal behavior between extroversion and introversion (Gallaher 1992)

- Communicative acts

According to the DIT++ taxonomy of dialogue acts (Bunt 2009), we did a preliminary analysis of communicative acts based on the "prepare dinner for friends" application scenario. We consider several relevant communicative acts to the scenario as summarized in Table 65.

The main communicative acts selected as relevant to the episodes in the scenario are: inform, confirm and warning as showed in Table 67.

General-purpose communicative functions	Examples of linguistic or non-verbal expressions	Use cases in the scenario
Information transfer functions (Information providing functions)	Inform: Speaker believes that the information he provides is correct.	The virtual character informs that given the current stove temperature, the pasta will boil within twelve minutes.
	Confirm: Speaker believes that Addressee believes that the propositional content is true.	The virtual character confirms that the pasta has been added.
	Warning: Speaker evaluates the situation, described in the semantic content, to be dangerous or potentially harmful for Addressee.	The virtual character warns that in a couple of minutes, the user will have to add the pasta. The virtual character warns that the cooking time is almost over and the stove

[5] The protocol of the video recording and the related scenario are provided in Annex A.

	will automatically switch off .

Table 67: General-purpose communicative functions and related expressions for the scenario "cooking pasta" (Bunt 2009)

The social obligation dimension of dimensions-specific functions describes the dialogue acts aiming at taking care of social conventions such as welcome, greetings, apologizes in case of mistakes or inability to help the dialogue partner, and farewell greetings. In the present scenario, two social obligation aspects - greeting and gratitude expressions – will be taken into account since they occurs much more frequently than any other expressions during the human communication. Greeting leads us for example to say "Hi" and raise our right arm, smiling, etc. Thanking is basic to most cultures for appreciation and recognition; it needs to be a set of actions that extends beyond only the verbal or use of words. In most cases, saying "thanks" is not enough, we accompany this word with smiling, nodding the head or even leaning forward the upper body. We design two different versions for each of the two cases according to the user experience with the system, as described in the column of use cases in the scenario in Table 67:

Table 68: Dimensions-specific functions and related expressions for the scenario "cooking pasta" (Bunt 2009)

dimensions-specific functions	Examples of linguistic or non-verbal expressions	Use cases in the scenario
Social obligations management functions	Salutation and Self-introduction	Greeting First time: in the first time the virtual character interacting with the user will give a self-introduction and then introduce the system. Later: the virtual character will greet the user if he hasn't seen him for more than 3 hours or just ask if everything is ok.
	Gratitude expressions	Thanking In between: at the end of the subtask, the virtual character will show that he is grateful for having done the job together with the user End: at the end of using experience, the virtual character would like to show that he is grateful for having shared experience with the user

F. Designing bodily expressions

Within the scope of our study, two modalities of bodily expressions are considered: gestures and postures. Previously, we selected what personality traits and what communicative acts would be performed. Another question rises up: how these personality trait and communicative acts can be communicated through these modalities? We suggested using a dual approach which is well suited for the design of virtual characters: literature and recording corpora. The procedure was briefly presented below and an example is given in Figure 55.



Figure 55: Given the same communicative act (inform cooking time), the agent exhibit different poses (from left to right): 1) the recorded subject; 2) the extravert agent; and 3) the introvert agent.

Step 1: A study of the literature in Behavioral Sciences and Psychology provided general results on personality, specific non verbal behavior and non-verbal lexicons for various communicative acts as summarized in Table 68. We suggested endowing the virtual character with the extraversion/introversion derived from the Big Five model of personality.

Step 2: A corpus will be collected and analyzed to make context dependent non-verbal behaviors of the virtual character. A set of scripts will be given to elicit relevant gestures and postures to the scenario. The corpus can be then analyzed.

Step 3: When performing the canned animations with Poser Pro ©, several parameters (as seen in Table 66) are under control to distinguish the bodily expressions of an extrovert virtual character from that of an introvert one. Then, we define a set of parameters which make it possible to perform distinctive gestures and postures. A summary of the specifications of the bodily expressions of the virtual character is presented in Table 69.

Table 69: A summary of designing bodily expressions

episodes	dialogue scripts	Descriptions of NVCs (postures and gestures)
salutation	hello!	Emblem_hands-up for extrovert, arms raise, happy hands; Emblem_hands-crossed for introvert.
inform_add_salt	It's time to add the salt.	Deictic_space, for the extravert, hand points into the space in front of the speaker and Akimbo (hands on hips) postures for both of the agents.
warn_add_pasta	In couple of minutes, you'll have to add the pasta.	Emblem-chide (warning) for extravert, open hand(s) or forefinger is shaken at somebody in an almost threatening manner (shaking). Emblem_attention for the introvert, forefingers raise (static).
confirm_add_pasta (*)	Yes, the pasta has been added.	Emblem-block for the extravert, hands are positioned in front of the speaker, palm toward addressee, in an italic movment to the side. An emblem gesture (well-done) can be occured before or after the confirmation gesture. Just Head-nodding for the introvert.
inform_cooking_time	So just to be patient, the pasta will come to boil in twelve minutes.	Meta_progress (a metaphor of change, evolution, and continuity) for the extravert, hands move in arc and one hand moves away from the body. For the extrovert, larger ampleur when the hand moves away.
warn_user_NV	Attention! The cooking time is almost over! (only nonverbal)	Emblem_chide (warning), open hand(s) or forefinger is shaken at somebody in an almost threatening manner; plus deictic_space
warn_user_V&NV	Attention! The cooking time is almost over! (verbal and nonverbal)	Emblem_attention, for both of the agents, hand points up. Akimbo (hands on hips) posture for the extravert.
gratitude inbetween*	well done!	Emblem_commended, hands in front of the check, while forefingers stand up.
gratitude end	It was nice to meet you. See you later!	Emblem_prayer, the palms of the two hands touch, finger tips pointing up, like in prayer; leaning forward posture

REFERENCES

Adolphs, R., D. Tranel, et al. (2003). "Dissociable neural systems for recognizing emotions." Brain and Cognition **52**(1): 61-69.

Alexandersson, J. Z., G., Carrasco, E (2010). D10.5: Publishable Final Activity Report, i2home: Intuitive Interaction for Everyone with Home Appliances based on Industry Standards.

Allwood, J. (1999). Cooperation and Flexibility in Multimodal Communication. 7th European Summer School on Language and Speech Communication, Stockholm, Sweden.

Argyle, M. (1975). Bodily communication. Second edition. London and New York, Routledge. Taylor & Francis.

Bailenson, J. N., J. Blascovich, et al. (2001). "Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments." Presence: Teleoper. Virtual Environ. **10**(6): 583-598.

Ballin, D., M.-F. Gillies, et al. (2004). A Framework For Interpersonal Attitude And Non-verbal Communication In Improvisational Visual Media Production. First European Conference on Visual Media Production IEE, London, UK

Bänziger, T., H. Pirker, et al. (2006). GEMEP - GEneva Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. Workshop "Corpora for research on emotion and affect". 5th International Conference on Language Resources and Evaluation (LREC'2006), Genova, Italy.

Barker, T. (2003). Collaborative Learning with Affective Artificial Study Companions in a Virtual Learning environment. Computer Based Learning Unit, The University of Leeds.

Barkhuysen, P., E. Krahmer, et al. (2010). "Crossmodal and incremental perception of audiovisual cues to emotional speech." Lang Speech **53**(1): 3-30.

Baron, R. M. and D. A. Kenny (1986). "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations." Journal of Personality and Social Psychology **51**(6): 1173-1182.

Battersby, S. A. (2011). Moving Together: The organisation of non-verbal cues during multiparty conversation. London, Queen Mary, University of London.

Bavelas, J. B., L. Coates, et al. (2000). "Listeners as co-narrators." Journal of Personality and Social Psychology **79**(6): 941-952.

Bellik, Y. (1995). Interfaces Multimodales : Concepts, Modèles et Architectures. Orsay, University Paris-South 11.

Bellik, Y., I. Rebaï, et al. (2009). Multimodal Interaction within Ambient Environments: An Exploratory Study. Human-Computer Interaction – INTERACT 2009**:** 89-92.

Bertrand, R., P. Blache, et al. (2008). "Le CID -Corpus of Interactional Data- : protocoles, conventions, annotations." Traitement Automatique des Langues **49**(3): 1-30.

Bianchi-Berthouze, N. and A. Kleinsmith (2003). "A categorical approach to affective gesture recognition." Connection Science **15**(4): 259-269.

Bilous, F. R. and R. M. Krauss (1988). "Dominance and accomodation in the conversational behaviours of same- and mixed gender dyads." Language & Communication **8**(3/4): 183-194.

Blascovich, J. (2002). Social influence within immersive virtual environments. The social life of avatars: presence and interaction in shared virtual environments, Springer-Verlag New York, Inc.**:** 127-145.

Bos, P., D. Reidsma, et al. (2006). Interacting with a Virtual Conductor

Entertainment Computing - ICEC 2006, Springer Berlin / Heidelberg. **4161:** 25-30.

Bressem, J. (2007). Recurrent form features in coverbal gestures. Third International Conf. of the International Society for Gesture Studies (ISGS).

Bucy, E. P. and C.-C. Tao (2007). "The Mediated Moderation Model of Interactivity." Media Psychology **9**(3): 647 - 672.

Bull, P. (1987). Posture and Gesture, Pergamon Press.

Bunt, H. (2009). The DIT++ taxonomy for functional dialogue markup. In Proceedings of the AAMAS 2009 Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts" (EDAML 2009), Budapest.

Butterworth, B. and G. Beattie (1978). Gesture and silence as indicators of planning in speech. Recent advances in the psychology of language: formal and experimental approaches. R. N. Campbell and P. Smith. New York, Plenum Press**:** 347-360.

Cacioppo, J. T., R. P. Petty, et al. (1986). "Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions." Journal of Personality and Social Psychology **50**: 260-268.

Caridakis, G., A. Raouzaiou, et al. (2006). Synthesizing Gesture Expressivity Based on Real Sequences. Workshop "Multimodal Corpora. From Multimodal Behaviour Theories to Usable Models". 5th International Conference on Language Resources and Evaluation (LREC'2006), Genova, Italy.

Carletta, J., Evert, S., Heid, U., and Kilgour, J. (2005). "The NITE XML Toolkit: data model and query." Language Resources and Evaluation Journal **39**(4): 313-334.

Cassell, J. (2007). Body Language: Lessons from the Near-Human. The Sistine Gap: History and Philosophy of Artificial Intelligence. J. Riskin, Chicago: University of Chicago Press.

Cassell, J., J. Sullivan, et al. (2000). Embodied Conversational Agents, MIT Press. .

Cassell, J., O. Torres, et al. (1999). Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation. Machine Conversations. Y. Wilks, The Hague: Kluwer**:** 143-154.

Cassell, J., O. E. Torres, et al. (1998). Turn taking vs. Discourse Structure: How Best to Model Multimodal Conversation. Machine Conversations, Kluwer.

Castellano, G. (2008). Movement Expressivity Analysis in Affective Computers: From Recognition to Expression of Emotion. Ph.D. Thesis, , , , . Department of Communication, Computer and System Sciences. Italy, University of Genoa.

Castellano, G., I. Leite, et al. (2010). "Affect recognition for interactive companions: challenges andÂ design in real world scenarios." Journal on Multimodal User Interfaces **3**(1): 89-98.

Castellano, G., A. Pereira, et al. (2009). Detecting user engagement with a robot companion using task and social interaction-based features. Proceedings of the 2009 international conference on Multimodal interfaces. Cambridge, Massachusetts, USA, ACM.

Cavicchio, F. and M. Poesio (2009). Multimodal corpora annotation: validation methods to assess coding scheme reliability. Multimodal corpora. K. Michael, M. Jean-Claude, P. Patrizia and H. Dirk, Springer-Verlag**:** 109-121.

Charny, E. J. (1966). "Psychosomatic manifestations of rapport in psychotherapy." Psychosomatic Medicine **28**: 305-315.

Chartrand, T. L. B., John A. (1999). "The chameleon effect: The perception–behavior link and social interaction." Journal of Personality and Social Psychology **76**(6): 893-910.

Chi, D., M. Costa, et al. (2000). The EMOTE model for effort and shape. Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co.

Clavel, C. and J.-C. Martin (2009). PERMUTATION: A Corpus-Based Approach for Modeling Personality and Multimodal Expression of Affects in Virtual Characters. HCI **11:** 211-220.

Clavel, C. and J.-C. Martin (2009). PERMUTATION: A Corpus-Based Approach for Modeling Personality and Multimodal Expression of Affects in Virtual Characters. . Digital Human Modeling Second International Conference, ICDHM 2009, Held as Part of HCI International 2009. V. G. Duffy. San Diego, Springer, LNCS 5620, 978-3-642-02808-3**:** 211-220.

Condon, W. S. and W. D. Osgton (1971). Speech and body motion synchrony of the speaker-hearer. The perceptionof Language. D. H. Hortonand and J. J. Jenkins, Academic Press**:** 150–184.

Core, M. G. and J. F. Allen (1997). Coding Dialogues with the DAMSL Annotation Scheme. AAAI Fall Symposium on Communicative Action in Humans and Machines, Menlo Park, California, American Association for Artificial Intelligence.

Coulson, M. (2004). "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence." Journal of Nonverbal Behavior **28**: 117-139.

Coulson, M. (2004). "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence." Journal of Nonverbal Behavior **28**(2): 117-139.

Courgeon, M., M.-A. Amorin, et al. (2010). Do Users Anticipate Emotion Dynamics in Facial Expressions of a Virtual Character? 23rd Annual Conference on Computer Animation and Social Agents (CASA 2010). Saint-Malo, France**:** Electronic proceedings.

Courgeon, M., J.-C. Martin, et al. (2008). MARC: a Multimodal Affective and Reactive Character. Workshop on Affective Interaction on Natural Environment, Chania, Greece.

Coutaz, J., J. L. Crowley, et al. (2005). "Context is key." Commun. ACM **48**(3): 49-53.

D'Mello, S., S. Craig, et al. (2008). "Automatic detection of learner's affect from conversational cues." User Modeling and User-Adapted Interaction **18**(1): 45-80.

D'Mello, S. and A. C. Graesser (2006). Affect Detection from Human-Computer Dialogue with an Intelligent Tutoring System. 6th International Conference on Intelligent Virtual Agents (IVA'06), Marina del Rey, CA, Springer.

Darwin, C. (1872). The expression of emotion in man and animal. Chicago, University of Chicago Press (reprinted in 1965).

Davis, F. D. (1989). "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." MIS Quarterly **13**(3): 319-339.

de Gelder, B., J. Snyder, et al. (2004). "Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body." Proceedings of the National Academy of Sciences **101**(47): 16701-16706.

de Gelder, B. and J. Van den Stock (2010). Real faces, real emotions: perceiving facial expressions in naturalistic contexts of voices, bodies and scenes. The handbook of face perception. G. R. A.J. Calder, J.V. Haxby & M.H. Johnson (Eds.), Oxford: Oxford University Press.

De Silva, P. R. and N. Bianchi-Berthouze (2004). "Modeling human affective postures: an information theoretic characterization of posture features." Computer Animation and Virtual Worlds **15**: 269-276.

Doucet, C. and R. M. Stelmack (1997). "Movement time differentiates extraverts from introverts." Personality and Individual Differences **23**(5): 775-786.

Dow, S., M. Mehta, et al. (2007). Presence and engagement in an interactive drama. Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, California, USA, ACM.

Ducatel, K., M. Bogdanowicz, et al. (2001). Scenarios for Ambient Intelligence European Commission. I. A. G. F. Report.

Edlund, J., J. Beskow, et al. (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. Proc. of LREC 2010, Valetta.

Egges, A., G. Papagiannakis, et al. (2007). "Presence and interaction in mixed reality environments." The Visual Computer **23**(5): 317-333.

Eissfeller, B., D. Gaensch, et al. (2004). Indoor Positioning Using Wireless LAN Radio Signals. IONGNSS Long Beach, California, USA**:** 1936-1947.

Ekman, P. (1964). "Body position, facial expression, and verbal behavior during interviews." The Journal of Abnormal and Social Psychology **68**(3): 295-301.

Ekman, P. (1965). "Differential communication of affect by head and body cues." Journal of Personality and Social Psychology **2**(5): 726-735.

Ekman, P., W. C. Friesen, et al. (2002). Facial Action Coding System. The Manual on CD ROM., Research Nexus division of Network Information Research Corporation.

Ekman, P. and W. V. Friesen (1967). "Head and body cues in the judgment of emotion: a reformulation." Perceptual And Motor Skills **24**(3): 711-724.

Ekman, P. and W. V. Friesen (1975). Unmasking the face. A guide to recognizing emotions from facial clues, Prentice-Hall Inc., Englewood Cliffs, N.J.

Fehr, B. and J. A. Russell (1984). "Concept of emotion viewed from a prototype perspective." Journal of Experimental Psychology **113**(3): 464-486.

Fogtmann, M. H., J. Fritsch, et al. (2008). Kinesthetic interaction: revealing the bodily potential in interaction design. Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat. Cairns, Australia, ACM.

Fontana, R. J., E. Richley, et al. (2003). Commercialization of an ultra wideband precision asset location system. IEEE Conference on Ultra Wideband Systems and Technologies**:** 369-373.

Frazier, P. A., A. P. Tix, et al. (2004). "Testing Moderator and Mediator Effects in Counseling Psychology Research." Journal of Counseling Psychology **51**(1): 115-134.

Frijda, N. H. (1987). "Emotion, cognitive structure, and action tendency." Cognition and Emotion **1**(2): 115-143.

Frijda, N. H., P. Kuipers, et al. (1989). "Relations among emotion, appraisal, and emotional action readiness." Journal of Personality and Social Psychology **57**(2): 212-228.

Gallaher, P. (1992). "Individual differences in nonverbal behavior: Dimensions of style." Journal of Personality and Social Psychology(63): 133–145.

Garrod, S. and M. J. Pickering (2004). "Why is conversation so easy?" Trends in Cognitive Sciences **8**(1): 8-11.

Giles, H., J. Coupland, et al. (1991). Contexts of accommodation: developments in applied sociolinguistics, Cambridge University Press.

Giles, H. and P. Johnson (1987). "Ethnolinguistic identity theory: a social psychological approach to language maintenance." International Journal of the Sociology of Language **1987**(68): 69-100.

Gillies, M., I. B. Crabtree, et al. (2006). Individuality and Contextual Variation of Character Behaviour for Interactive Narrative. AISB Workshop on Narrative AI and Games

Goodwin, M. H. and C. Goodwin (2000). Emotion within Situated Activity. . Linguistic Anthropology: A Reader. MA, Oxford**:** 239-257

Gratch, J., A. Okhmatovskaia, et al. (2006). Virtual Rapport. 6th International Conference on Intelligent Virtual Agents (IVA'06), Marina del Rey, CA, Springer.

Gross, J. J. and L. F. Barrett (2011). "Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View." Emotion Review **3**(8).

Hall, E. T. (1963). "A System for the Notation of Proxemic Behaviour." American Anthropologist 1003-1026. .

Hall, E. T. (1966). The Hidden Dimension: Man's Use of Space in Public and Private. London, UK, The Bodley Head Ltd.

Harrigan, J. A., R. Rosenthal, et al. (2005). The new handbook of methods in nonverbal behavior research, Oxford University Press.

Harrigan, J. A., R. Rosenthal, et al. (2005). The new handbook of methods in nonverbal behavior research. Oxford.

Hartmann, B., M. Mancini, et al. (2005). Design and Evaluation of Expressive Gesture Synthesis for Embodied Conversational Agents. AAMAS'05.

Holler, J. (2007). The influence of interactional processes on speakers' use of gesture space. Deafness, Cognition & Language Research Centre (DCAL). London.

Hooijdonk, C. M. J. v. (2008). Explorations in multimodal information presentation. Enschede, PrintPartners Ipskamp.

Huang, L., L.-P. Morency, et al. (2011). Virtual Rapport 2.0. Intelligent Virtual Agents, Springer Berlin / Heidelberg.

Ibister, K. and C. Nass (2000). "Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics." Int. J. Human-Computer Studies(53): 251-267.

Jannedy, S. and N. Mendoza-Denton (2005). "Structuring nformation through gesture and intonation." Interdisciplinary Studies on Information Structure **3**: 199-244.

Kasap, M. and N. Magnenat-Thalmann (2009). Sizing avatars from skin weights. 16th ACM Symposium on Virtual Reality Software and Technology, Kyoto, Japan, ACM.

Kaufman, L. (1974). Sight and mind: an introduction to visual perception. New York, Oxford University Press.

Kendon, A. (1972). Some relationships betwen body motion and speech. Studies in Dyadic Communication. A. Siegman and B. Pope. New York, Pergamon Press**:** 177-210.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. The relationship of Verbal and Nonverbal Communication. M. R. Key, Mouton Publishers**:** 207-228.

Kendon, A. (2004). Gesture: Visible Action as Utterance. Cambridge, Cambridge University Press.

Kendon, A., A. Esposito, et al. (2010). Spacing and Orientation in Co-present Interaction. Development of Multimodal Interfaces: Active Listening and Synchrony, Springer Berlin / Heidelberg. **5967:** 1-15.

Kimbara, I. (2006). "On gestural mimicry." Gesture. John Benjamins Publishing Company. **6**(1): 39-61.

Kipp, M. (2001). Analyzing individual nonverbal behavior for synthetic character animation. Colloque Oralité et Gestualité (ORAGE 2001), Paris: L'Harmattan.

Kipp, M. (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. 7th European Conference on Speech Communication and Technology (Eurospeech'2001), Aalborg, Danemark.

Kipp, M. (2004). Gesture Generation by Imitation. From Human Behavior to Computer Character Animation. Florida, Boca Raton, Dissertation.com.

Kipp, M., J. C. Martin, et al., Eds. (2009). Multimodal Corpora: From Models of Natural Interaction to Systems and Applications. Lecture Notes on Artificial Intelligence, LNAI 5509, Springer.

Kipp, M., M. Neff, et al. (2006). An Annotation Scheme for Conversational Gestures: How to economically capture timing and form. Workshop "Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models " In Association with the 5th International Conference on Language Resources and Evaluation (LREC2006), Genoa, Italy.

Kita, S., I. van Gijn, et al. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. Gesture and Sign Language in Human Computer Interaction: Proceedings / International Gesture Workshop, Bielefeld, Germany, Springer-Verlag: Berlin Heidelberg.

Kleinsmith, A. and N. Bianchi-Berthouze (2007). Recognizing affective dimensions from body posture. 2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007), Lisbon, Portugal, Springer, LNCS, vol. 4738.

Kleinsmith, A., N. Bianchi-Berthouze, et al. (2010). Form as a Cue in the Automatic Recognition of Non-acted Affective Body Expressions

Affective Computing and Intelligent Interaction, Springer Berlin / Heidelberg. **6974:** 155-164.

Knapp, M. L. and J. Hall (2006). Nonverbal Communication in Human Interaction.

Krauss, R. M. (1998). "Why Do We Gesture When We Speak?" Current Directions in Psychological Science **7**: 54-59.

Kret, M. E., S. Pichon, et al. (2011). "Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study." NeuroImage **54**(2): 1755-1762.

Kruijff-Korbayov, I., C. G. O. Kukina, et al. (2008). Generation of output style variation in the SAMMIE dialogue system. Proceedings of the Fifth International Natural Language Generation Conference. Salt Fork, Ohio, Association for Computational Linguistics.

Kruppa, M., L. Spassova, et al. (2005). The virtual room inhabitant – intuitive interaction with intelligent environments. Proceedings of the 18th Australian Joint Conference on Artificial Intelligence (AI05), Sydney, Australia.

Kuhn, S., B. C. N. Muller, et al. (2010). "Why do I like you when you behave like me? Neural mechanisms mediating positive consequences of observing someone being imitated." Social Neuroscience **5**(4): 384-392.

LaFrance, M. (1979). "Nonverbal synchrony and rapport: analysis by the cross-lag panel technique." Social Psychology Quarterly **42**: 66-70.

LaFrance, M. and M. Broadbent (1976). "Group Rapport: Posture Sharing as a Nonverbal Indicator." Group and Organizational Studies **1**: 328-333.

LaFrance, M. and W. Ickes (1981). "Psoture mirroring and interactional involvement:sex and sex-typing effects." Journal of Nonverbal Behaviour **5**: 139-154.

Lakin, J. L., V. E. Jefferis, et al. (2003). "The Chameleon Effect as Social Glue: Evidence for the Evolutionary Significance of Nonconscious Mimicry." Journal of Nonverbal Behavior **27**(3): 145-162.

Lazarus, R. S. (1991). "Progress on a cognitive-motivational-relational theory of emotion." American Psychologist **46**(8): 819-834.

Leppanen, J. M. and J. K. Hietanen (2004). "Positive facial expressions are recognized faster than negative facial expressions, but why?" Psychological Research **69**(1): 22-29.

Löckelt, M., T. Becker, et al. (2002). Making Sense of Partial. Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), Edinburgh

Loehr, D. (2004). Gesture And Intonation. Faculty of the Graduate School of Arts and Sciences of Georgetown University.

Lohse, M. (2009). Investigating the influence of situations and expectations on user behavior - empirical analyses in human-robot interaction, Technische Fakultät Universität Bielefeld.

Mancini, M. and C. Pelachaud (2007). Dynamic Behavior Qualifiers for Conversational Agents. Proceedings of the 7th international conference on Intelligent Virtual Agents. Paris, France, Springer-Verlag.

Marsella, S., J. Gratch, et al. (2010). Computational Models of Emotion. Blueprint for Affective Computing. K. R. Scherer, T. Banziger and E. Roesch, Oxford University Press**:** 21-41.

Mayer, R. E. (2002). "Multimedia learning." Psychology of Learning and Motivation **41**: 85-139.

McNeill, D. (1992). Hand and mind - what gestures reveal about thoughts, University of Chicago Press, IL.

McNeill, D. (2005). Gesture and Thought, The University of Chicago Press.

McNeill, D. (2005). Gesture and thought: codification and social interaction, Embodied Communication I. Opening Conference of the ZiF: Research Group 2005/2006 "Embodied Communication in Humans and Machines". Scientific organization: Ipke Wachsmuth (Bielefeld), Günther Knoblich (Newark).

McNeill, D. and S. D. Duncan (2000). Growth points in thinking-for-speaking. Language and Gesture. D. McNeill, Cambridge: Cambridge University Press**:** 141-161.

McNeill, D., F. Quek, et al. (2001). "Catchments, prosody and discourse." Gesture **1**(1): 9-33.

McQuiggan, S. W., J. P. Rowe, et al. (2008). The effects of empathetic virtual characters on presence in narrative-centered learning environments. twenty-sixth annual SIGCHI conference on Human factors in computing systems. Florence, Italy, ACM.

Meillon, B., A. Tcherkassof, et al. (2010). DYNEMO: A Corpus of dynamic and spontaneous emotional facial expressions LREC. Valleta.

Miyawaki, K. and M. Sano (2008). A virtual agent for a cooking navigation system using augmented reality. 8th international Conference on intelligent Virtual Agents Tokyo, Japan.

Mol, L., E. Krahmer, et al. (2011). "Adaptation in gesture: Converging hands or converging minds?" Journal of Memory and Language.

Molina, F. J. V. (2010). Ambient Intelligence InTech

Montoro, G., P. Haya, et al. (2008). A Study of the Use of a Virtual Agent in an Ambient Intelligence Environment. Intelligent Virtual Agents**:** 520-521.

Nakano, Y. I. and R. Ishii (2011). Estimating user's engagement from eye-gaze behaviors in human-agent conversations. Proceedings of the 15th international conference on Intelligent user interfaces, Hong Kong, China, ACM.

Neff, M. and E. Fiume (2006). "Methods for Exploring Expressive Stance." Graphical Models.Special issue on SCA 2004. **68** (2): 133 - 157.

Newlove, J. (1993). Laban for actors and dancers. New York, Routledge.

Nixon, M., P. Pasquier, et al. (2010). DelsArtMap: Applying Delsarts Aesthetic System to Virtual Agents. Intelligent Virtual Agents, Springer Berlin / Heidelberg.

Oviatt, S. (1996). Multimodal interfaces for dynamic interactive maps. Proceedings of the SIGCHI conference on Human factors in computing systems: common ground. Vancouver, British Columbia, Canada, ACM.

Oviatt, S. (2002). Multimodal Interfaces. Handbook of Human-Computer Interaction. J. J. A. Sears. New Jersey, Lawrence Erlbaum.

Özyürek, A. (2002). "Do Speakers Design Their Cospeech Gestures for Their Addressees? The Effects of Addressee Location on Representational Gestures." Journal of Memory and Language **46**(4): 688-704.

Pantic, M. and I. Patras (2006). "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences." IEEE Trans Syst Man Cybern B Cybern. **36**(2): 433-49.

Park, N., K. M. Lee, et al. (2010). "Effects of pre-game stories on feelings of presence and evaluation of computer games." Int. J. Hum.-Comput. Stud. **68**(11): 822-833.

Parker, S. P. (2003). McGraw-Hill dictionary of scientific and technical terms, McGraw-Hill.

Pascal, B. and R. W. Gleason (1966). The essential Pascal, New American Library.

Pedica, C., H. Hogni, et al. (2009). Spontaneous Avatar Behavior for Human Territoriality. Proceedings of the 9th International Conference on Intelligent Virtual Agents. Amsterdam, The Netherlands, Springer-Verlag.

Pelachaud, C. (2005). Multimodal expressive embodied conversational agent. ACM Multimedia, Brave New Topics session. Singapore, ACM**:** 683 - 689.

Peters, C., C. Pelachaud, et al. (2005). A model of attention and interest using Gaze behavior. Lecture Notes in Computer Science, Springer-Verlag**:** 229-240.

Pianesi, F., M. Zancanaro, et al. (2006). Multimodal Annotated Corpora of Consensus Decision Making Meetings Workshop "Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models " in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006), Genoa, Italy.

Picard, R. (1997). Affective Computing, MIT Press.

Picard, S. M. a. R. W. (2003). Automated Posture Analysis for Detecting Learner's Interest Level. Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction CVPR HCI.

Pineda, L. A., A. Massa, et al. (2002). The DIME Project. MICAI 2002 : Advances in Artificial Intelligence, Springer Berlin / Heidelberg.

Pitterman, H. and S. Nowicki Jr (2004). "A Test of the Ability to Identify Emotion in Human Standing and Sitting Postures: The Diagnostic Analysis of Nonverbal Accuracy-2 Posture Test (DANVA2-POS)." Genetic, Social & General Psychology Monographs **130**: 146-162.

Platzer, W. and K. W. (2004). Color Atlas and Textbook of Human Anatomy, Thieme.

Poggi, I. and C. Pelachaud (2000). Performative facial expressions in animated faces. Embodied Conversational Agents. J. Cassell, S. Prevost and E. Churchill, MIT-Press**:** 155-188.

Poggi, I., C. Pelachaud, et al. (2000). "Eye communication in a conversational 3D synthetic agent." AI Commun. **13**(3): 169-181.

Pousman, Z. and J. Stasko (2006). A taxonomy of ambient information systems: four patterns of design. Proceedings of the working conference on Advanced visual interfaces, Venezia, Italy, ACM.

Quek, F., D. McNeill, et al. (2002). "Multimodal human discourse: gesture and speech." ACM Transactions on Computer-Human Interaction **9**(3): 171-193.

Rao, S. P. and D. J. Cook (2004). "Predicting inhabitant action using action and task models with application to smart homes " International Journal on Artificial Intelligence Tools **13**(1): 81-99.

Raymond, C., K. J. Rodríguez, et al. (2008). Active Annotation in the LUNA Italian Corpus of Spontaneous Dialogues. Proceedings of the sixth international conference on Language Resources and Evaluation (LREC 2008), Marrakech. Marrocco.

Reed, C., J. Garza, et al. (2007). The Influence of the Body and Action on Spatial Attention. Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint**:** 42-58.

Reeves, B. and C. Nass (2000). "Perceptual user interfaces: perceptual bandwidth." Commun. ACM **43**(3): 65-70.

Reeves, B., Thorson, E., Rothschild, M. L., McDonald, D., Hirsch, J., Goldstein, R. (1985). "Attention to television: intrastimulus effects of movement and scene changes on alpha variation over time." International Journal of Neuroscience **27**(3-4): 241–255.

Rehm, M. and E. André (2008). From Annotated Multimodal Corpora to Simulated Human-Like Behaviors. Modeling Communication with Robots and Virtual Humans. I. W. a. G. Knoblich, Berlin: Springer**:** 1-17.

Rehm, M., E. André, et al. (2005). Let's Come Together - Social Navigation Behaviors of Virtual and Real Humans. Berlin, Heidelberg.

Rehm, R., E. André, et al. (2005). Let's Come Together - Social Navigation Behaviors of Virtual and Real Humans. INTETAIN 2005 Springer, Berlin, Heidelberg.

Reidsma, D., D. Heylen, et al. (2006). Annotating Emotion in Meetings. 5th international conference on Language Resources and Evaluation (LREC 2006). Genoa, Italy.

Richmond, V. P. and J. C. Croskey (1999). Non Verbal Behavior in Interpersonal relations, Allyn & Bacon Inc.

Rist, T. and E. Andre (1993). Designing Coherent Multimedia Presentations. Proceedings of the Fifth International Conference on Human-Computer Interaction 1993.

Rist, T. and E. Andre (2003). Building smart embodied virtual characters. Proceedings of the 3rd international conference on Smart graphics. Heidelberg, Germany, Springer-Verlag.

Rodrigues, I., S. Kopp, et al. Gesture Space and Gesture Choreography in European Portuguese and African Portuguese Interactions: A Pilot Study of Two Cases

Gesture in Embodied Communication and Human-Computer Interaction, Springer Berlin / Heidelberg. **5934:** 23-33.

Rousseau, C. (2006). Présentation multimodale et contextuelle de l'information. Computer Sciences. Orsay, University of Paris-Sud.

Roussel, N., H. Evans, et al. (2004). "Proximity as an Interface for Video Communication." IEEE MultiMedia **11**(3): 12-16.

Russell, J. A. (1980). "A circumplex model of affect." Journal of Personality and Social Psychology **39**(6): 1161-1178.

Ruttkay, Z. and C. Pelachaud, Eds. (2004). From brows to trust: evaluating embodied conversational agents, Kluwer Academic Publishers.

Ruttkay, Z. f., J. Zwiers, et al. (2006). Towards a Reactive Virtual Trainer. Intelligent Virtual Agents, Springer Berlin / Heidelberg. **4133:** 292-303.

Sacks, H., E. A. Schegloff, et al. (1974). "A simplest systematics for the organization of turntaking for conversation." Language **50**: 696-735.

Sanghvi, J., G. Castellano, et al. (2010). Automatic analysis of affective postures and body motion to detect engagement with a game companion. Proceedings of the 6th international conference on Human-robot interaction. Lausanne, Switzerland, ACM.

Scheflen, A. E. (1964). "The significance of posture in communication systems." Psychiatry **27**: 316-331.

Scherer, K. R. (2000). Emotion. Introduction to Social Psychology: A European perspective. M. H. W. Stroebe, Oxford: Blackwell**:** 151-191.

Scherer, K. R. (2010). The component process model: Achitecture for a comprehensive computational model of emergent emotion. Blueprint for Affective Computing. K. R. Scherer, T. Banziger and E. Roesch, Oxford University Press**:** 47-70.

Scherer, K. R. (2010). Emotion and emotional competence: conceptual and theoretical issues for modelling agents. Blueprint for Affective Computing. K. R. Scherer, T. Banziger and E. Roesch, Oxford University Press.

Scherer, K. R., A. Schorr, et al. (2001). Appraisal considered as a process of multilevel sequential checking. Appraisal processes in emotion: Theory, methods, research. New York, NY US, Oxford University Press**:** 92-120.

Sidner, C. L. and C. Lee (2007). Attentional Gestures in Dialogues Between People and Robots. Conversational Informatics, John Wiley & Sons, Ltd**:** 103-115.

Smith, P. K. (2008). "Non conscious effects of power on basic approach and avoidance locomotion." Social Cognition **26**(1): 1-24.

Spiekermann, S. (2007). User Control in Ubiquitous Computing: Design Alternatives and User Acceptance. Institut für Wirtschaftsinformatik. Berlin, Humboldt-Universitat zu Berlin. **Habiliation**.

Stanchfield, W. and D. Hahn (2009). Drawn to Life: 20 Golden Years of Disney Master Classes: Volume 2: The Walt Stanchfield Lectures, Elsevier Science & Technology.

Steggles, P. and S. Gschwind (2005). The Ubisense smart space platform. Third International Conference on Pervasive Computing.

Sweetser, E. and M. Sizemore (2008). Personal and interpersonal gesture spaces: Functional contrasts in language and gesture. Language in the Context of Use, Mouton de Gruyter. **37:** 25-52.

Sweetser, E. and M. Sizemore (2008). Personal and interpersonal gesture spaces: Functional contrasts in language and gesture. Language in the Context of Use, Mouton de Gruyter. **Volume 37:** 25-52.

Tan, H. Z., L. A. Slivovsky, et al. (2001). "A sensing chair using pressure distribution sensors." Mechatronics, IEEE/ASME Transactions on **6**(3).

Tan, N., G. Pruvost, et al. (2010). A location-aware virtual character in a smart room: effects on performance, presence and adaptivity. Proceedings of the 16th international conference on Intelligent user interfaces. Palo Alto, CA, USA, ACM.

Tcherkassof, A. (2008). Les Emotions et leurs expressions.

Termine, N. I., C. (1988). "Infants' reaction to their mothers' expressions of joy and sadness." Developmental Psychology **24**: 223-229.

Thomas, F. and O. Johnston (1995). The illusion of life: Disney animation, Hyperion.

Trout, D. L. and H. M. Rosenfeld (1980). "The effect of postural lean and body congruence on the judgment of psychotherapeutic rapport." Journal of Nonverbal Behavior **4**(3): 176-190.

van Baaren, R. B., R. W. Holland, et al. (2003). "Mimicry for money: Behavioral consequences of imitation." Journal of Experimental Social Psychology **39**(4): 393-398.

Vanhala, T., V. Surakka, et al. (2010). "Virtual proximity and facial expressions of computer agents regulate human emotions and attention." Computer Animation and Virtual Worlds **21**(3-4).

Venkatesh, V. and F. D. Davis (2000). "A theoretical extension of the technology acceptance model: four longitudinal field studies." Manage. Sci. **46**(2): 186-204.

Wallbott, H. G. (1998). "Bodily expression of emotion." European Journal of Social Psychology **28**(6): 879-896.

Wallbott, H. G. (1998). "Bodily expression of emotion." European Journal of Social Psychology **28**: 879-896.

White, T. D., Folkens, P. A. (1991). Human Osteology. San Diego, Academic Press, Inc.

Wiendl, V., K. D. Ulhaas, et al. (2007). Integrating a Virtual Agent into the Real World: The Virtual Anatomy Assistant Ritchie. Intelligent Virtual Agents. Paris, France, Springer-Verlag.

Witmer, B. G. and M. J. Singer (1998). "Measuring Presence in Virtual Environments: A Presence Questionnaire." Presence: Teleoper. Virtual Environ. **7**(3): 225-240.

Yabar, Y., L. Johnston, et al. (2006). "Implicit Behavioral Mimicry: Investigating the Impact of Group Membership." Journal of Nonverbal Behavior **30**(3): 97-113.

Young, R. D., & Frye, M. (1966). "Some are laughing; some are not : why?" Pychological Reports **18**: 747-752.

Zhao, L. (2001). Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures, CIS, University of Pennsylvania.