CROSS CULTURAL COMPUTER-SUPPORTED COLLABORATION

BY

GRÉGOIRE A. LANDEL

BACHELOR OF SCIENCE IN ENGINEERING
CIVIL ENGINEERING AND OPERATIONS RESEARCH
JUNE 1998
PRINCETON UNIVERSITY

SUBMITTED TO THE DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUNE 1999

SIGNATURE OF
AUTHOR_____
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
May 14, 1999

CERTIFIED
BY_____
FENIOSKY PEÑA-MORA
Assistant Professor, Department of Civil and Environmental Engineering
Thesis Supervisor

APPROVED
BY_____
ANDREW J. WHITTLE
Chairman, Departmental Committee on Graduate Studies

Eng

CROSS CULTURAL COMPUTER-SUPPORTED COLLABORATION

BY

GRÉGOIRE A. LANDEL

SUBMITTED TO THE DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING ON MAY 14, 1999
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL ENGINEERING

ABSTRACT

The present thesis introduces the need for computer supported collaboration tools to provide nonverbal behavior information in order to foster effective communications within organizations. Effective communication can help companies achieve their goals by making sure that these goals are properly understood and agreed upon by all involved.

Since nonverbal behavior is crucial to communication, and since numerous companies operate across cultural and geographical borders, it is suggested that a Nonverbal Behavior Cultural Translator (NBCT) be implemented. After providing a definition of culture specific to the needs of the (NBCT), the thesis describes some methods for capturing, interpreting, translating, and representing nonverbal behavior.

This thesis does not purport to provide an end-all solution to cross cultural collaboration problems, but rather to create a framework for considering such problems. In particular, for considering how computer supported solutions might be useful in solving this daily problem of international companies.

THESIS SUPERVISOR: FENIOSKY PEÑA MORA
TITLE: ASSISTANT PROFESSOR OF CIVIL AND ENVIRONMENTAL ENGINEERING

Acknowledgments

I would like to take this opportunity to thank the people who have made it possible for me to write this thesis.

Professor Feniosky Peña-Mora, my thesis advisor, for his astute help and comments.

Michel, Marie Yvonne, Morgane and Bertrand Landel, my family, without whose love and support I would never have made it this far.

Professor Raphael Bras, Dr. Eric Adams, and the **MIT Civil Engineering Dept.**, who made it possible for me to attend MIT both by supporting me in my educational needs and by granting me a half tuition fellowship.

Professor James A. Smith, my undergraduate thesis advisor at Princeton University, who showed me how academic research is done.

Jacob Rasmussen, my roommate, and **Padmanabha Vedam,** a fellow student, who answered my questions relating to linear algebra

Jason Williams, Robert Jensen, Dr. Alain Mignault, Pofessor Rosalind Picard, and **Hannes Vilhjálmsson**, who provided insights and ideas for this thesis.

Table of Contents

List of Figures:

Chapter 1 Introduction and Motivations

O you who want to circle the Earth, listen to this pleasant fable
And for a long journey do not depart in haste
For no matter what your imagination might picture
There is no sweeter country than that were your friend and beloved reside

Ivan Kryvlov

The overall goal of this thesis is to present both the need and a method for translating culturally specific nonverbal behavior. The work will focus on applications to complement Computer Supported Collaborative Work tools. As will be demonstrated in this first chapter, there is a need for both a set of culturally sensitive tools for better communication than are at present available to companies who attempt to do business across cultural boundaries.

This is why the idea of a nonverbal behavior cultural translator (NBCT) arose. This chapter will also briefly describe the requirements of such a translator. In later chapters, this document will discuss the specific implementation challenges and design options which will need to be overcome to implement the nonverbal behavior cultural translator. A brief specific outline of the contents of the paper will be given at the end of this chapter.

This thesis is not about to deliver a final solution to the problem of computer supported cross cultural collaborative work. Rather, it is meant to provide a framework for thinking about issues related to this topic. The solutions provided here are first order approximations, of a final solution. The author believes that individual technologies required for implementing a NBCT exist, but that much work will yet be needed to find the most effective way to combine these technologies in the most efficient manner.

Section 1.1 CSCW is Becoming More Common

Because companies and organizations are becoming increasingly distributed, they are augmenting their use of Computer Supported Collaborative Work (CSCW) tools to supplement or replace face to face contact or collocated collaboration (Mark, 1998). The goal is to reduce the costs of collaborating with clients, colleagues, or suppliers by allowing companies to limit the travel formerly required of individuals who need to collaborate with physically distant partners, and by thus reducing the time needed by a team to accomplish a task. By leveraging computer technologies, and allowing these same individuals to collaborate over computer networks, companies are able to start to fulfill the needs of virtual teams (Mark, 1998). Of course, the telephone already fills part of the need for distributed collaboration, but it is limited in its capabilities. For example, it does not allow files (e.g.: pictures or text) to be easily shared, and it can only carry voice signals.

The plethora of applications and research projects sponsored by corporations which are devoted to allowing for and improving CSCW is a testament to the rising importance of this type of collaboration in the workplace. In fact, several companies are actively developing virtual workplaces for their distributed team to interact, or researching how to establish such virtual workplaces (e.g.: Lucent (Boyer et al., 1998), Boeing (Fuchs et al., 1998, Mark, 1998), Fuji Xerox (Adams and Toomey, 1998), and British Telecom (McGrath, 1998)).

While virtual communities mostly emerged from university settings into purely ludic beginnings supported by the world wide web (e.g.: chat rooms), the above companies' current goal of establishing virtual communities is to foster real work between distributed coworkers. Specifically, companies generally seek to allow their employees to achieve four different goals within virtual work environments, namely (Adams and Toomey, 1998):

- Work,
- Preserve organizational memory,
- Promote corporate culture, and
- Allow for professional networking.

In this thesis, the focus will be on professional networking, and on how virtual work environments can be improved to allow for better interaction between people both at a personal level and at a professional level. Professional networking involves interacting both socially and professionally with colleagues, clients, or members of the same industry in order to establish ties which might be used later in strictly business

settings. The next section focuses on what is meant by communications, and what human beings need in order to communicate effectively.

Section 1.2 Challenges in Developing Effective Communications

1.2.1 Effective Communication

Mantovani (1996) described how achieving a state in which communications can occur completely unhindered is conditioned by reaching a "shared symbolic order" and a shared definition of "appropriateness" for individual events or phrases. This state of communication will be referred to as "effective communication" in this thesis. "Shared symbolic order" means that communicative events have a single meaning for two communicating parties, whereas appropriateness is used to describe the limits of acceptable behavior within a relationship or of a group of people to which both communicating parties belong. In other words, only once there is "reciprocity" in a relationship will it be productive in the workplace, when the "expectations of a given behavior or response are the same for a given triggering action" (Mantovani, 1996).

While such a high level of communicative effectiveness between two people seems entirely unattainable in practice, it should be understood that, in general, individuals can understand each other completely about certain subjects, and not at all about others. In this case, reaching a good communicative state, especially one limited in scope, is possible (and often achieved, according to Mantovani (1996)). For people who

have no prior knowledge of each other, it will be necessary to build up to such a common understanding. Building a shared symbolic order, agreeing on appropriateness, and reaching reciprocity require personal contact at several levels. Indeed, to understand each other, two will need to know about each other, if only within the limited scope to which they might wish to confine their mutual understanding, and so, in a very loose sense, will need to communicate, or exchange information. As will be explained below, this communication can occur easily when individuals are collocated, but it is more difficult when they are not in each other's immediate presence because of the specific communication challenges which result from working in distributed environments.

The relationship between the process of socialization and the process of reaching communicative effectiveness will be discussed first. The specific challenges which result from working in distributed environments will also be addressed. Finally, there will be a brief discussion of the difficulties encountered in cross cultural communications.

1.2.2 Trust and Effective Communication

As explained in the previous section, the development of effective communication between two people is inherently linked to attaining a better understanding of each other's personal symbolic order, and standards of appropriateness. Of course, one has to learn the other's behavioral rules before a common set can be agreed on and reciprocity can be established. This is because individuals are constrained by their rituals, or rules of behavior (Mantovani (1996), Labarre, 1947). Moreover, one needs to be convinced that

the other's rituals are worthy of respect and trust if they are to be included in a shared symbolic order. Thus, the development of effective communication between two people is both dependent, and similar to the development of trust between these same individuals.

There are three types of trusts in professional relationships. The first is deterrence-based trust (Tyler and Kramer, 1996), which is based on the belief that punishment for failure to collaborate will be too high to justify a breech of trust. The second type of trust is called knowledge-based trust, and it is based on "knowing the other sufficiently well so that the other's behavior is predictable" (Lewicki and Bunker, 1996). Finally, identification-based trust occurs when one of the parties identifies with the other's goals and intentions (Lewicki and Bunker, 1996). The parallels between the development of effective communication and trust are interesting, since they most likely feedback from each other. Indeed, it is unlikely that knowledge-based trust can occur between two people who are not communicating, but such trust also seems to be the prerequisite for the very reciprocity which is necessary to engage in effective communications.

Lewicki and Bunker (1996) argue that the development of trust can is dependent on the experience of trustable behavior from the other party, through socialization and interaction. This experience can be gathered in two manners. First directly, through first-hand interaction with the other person an individual can come to witness the trustworthiness of the other (or lack thereof). Second vicariously, through the use of a

"Web of Trust" concept, an individual will rely on a trusted other's assurance to decide to trust a third party. This occurs commonly between humans, and can also be called a referral process. Referrals allow people to trust each other through the vouching of a third party. This is method of trust development is used most often for job and school applications.

Despite the two options available, the development of trust is a watching game, in which the limits, strengths, and weaknesses of the other must be integrated into a decision to trust or not to trust. Clearly, socialization, both professional and non professional, and contact between the two parties is determinant in this exercise. But as was seen above, it is only after trust is established that the rituals sharing which is the prerequisite of effective communication can take place. Therefore, socialization is required for the emergence of both trust and effective communication, and it should be fostered by the communication environment provided to co-workers.

McGrath (1998) recognized the need for socialization in work environments, both virtual and not. He therefore decided to include a space for "hanging out" in his "Forum". He wanted to ensure that proper relationships could be developed in this virtual work space, and he thus implicitly recognized the need for social interaction to foster both trust and effective communication. As the Forum is still a prototype, no specific results have been produced yet, but McGrath (1998) seems convinced that social activities play an important part of developing professional relationships. Adams and Toomey (1998) observed first hand the need for social interaction being expressed by

distributed team members working at Fuji Xerox, as individuals spontaneously engaged in nonprofessional conversations and shared personal information with their coworkers in the distributed environments which they shared.

Section 1.3 Interaction, Communication, and the Importance of Nonverbal Behavior

1.3.1 Socialization

Socialization is loosely defined as a set of interactions between individuals which occur at a personal level. The success of these interactions for any two people is controlled by several factors, including their ability to understand each other at the spoken language level (Adams and Toomey, 1998); indeed, people must be able to engage in conversation before they can reach an understanding and learn to communicate more effectively. Picard and Cosier (1997) explain that just as communication (exchanging information) are an important part of socialization, so are key elements of socialization (including exchanging affective information (Picard, 1995) central to communications. How people come to engage in conversations is a research topic unto itself, and the issue will not be discussed here. Rather, the focus will be on identifying those elements of communication (or information exchange) which make it effective.

1.3.2 Multi-Channel Communications

Communication specialists like Ray Birdwhistell (1970) and David McNeill (in Goldin-Weaver (1997)) have long argued that communication is not, and should not be considered to be limited to spoken or written exchange alone. In fact, Birdwhistell (1970) argues that human communication occurs using several parallel channels, which are all important in exchanging information. These channels make use of all of the available human senses, and at least one channel is always in use. It is important to realize that spoken information is only a part of the communicative experience, and by no means always the most important one. In an exchange, one party could be using speech as her primary mode of information output, while the other would be providing feedback using gestures as his primary mode of output.

A single sensory channels can contain several information channels simultaneously, e.g.: the content of sound information (all information gathered by hearing) is characterized both by the meaning of words and by the vocal modulations of pitch and tone used in producing the words. It would be wrong and overly simplistic to consider communication as a "verbal process" modified by "gestures, pushing and holding, tasting, and odor emitting or receiving". Rather, communication is more than the sum of its individual parts (Birdwhistell, 1970), and gesture in particular is wholly integrated and interwoven with speech (McNeill in Goldin-Weaver (1997)).

The direct consequence of the multi-channel nature of information exchange is that face to face communication is a very rich type of exchange, especially as compared

with written information exchange (Picard and Cosier (1997), Fridlund (1994), Picard (1995)). The participants are bombarded with inputs from each other, which they must quickly process within the context of the conversation in order to understand each other, at least at a superficial level. In fact, the rich content of information exchange is not a function of whether or not effective communication has not been established. Instead, it can be said that effective communication is established only when all (or most) information provided by the parties is mutually understood; this is another definition for reciprocity. Rapid processing of multi-channel information also allows the conversation to be interactive, in that the interlocutor can respond to the speaker in real time, without interrupting the speaker (Birdwhistell, 1970). For example, the listener can avoid interrupting the speaker by "acting": "I understand, and I am listening", instead of having to explicitly say it. Thus, two channels are in use, and information can go both ways simultaneously between speaker and listener.

Despite Birdwhistell's insistence that all sensory channels are at least theoretically important in communication, the perceived particular importance of gestures and nonverbal behavior in information exchange has spawned numerous studies (Picard (1995), Picard and Cosier (1997)). Psychologists such as Paul Ekman have extensively studied, classified, and analyzed nonverbal behavior in general, and others have determined which gestures where specialized for dialogue, and how they contribute specifically to the exchange of information between people (Bavelas et al. (1995), Goldin-Meadow (1997), Walther (1995)). The importance of nonverbal behavior within dialogue is twofold; it helps both the speaker and interlocutor. First, as explained above,

the interlocutor can provide silent, real time feedback to the speaker (e.g.: disagreeing, agreeing, being bored, paying attention) without speaking, simply by engaging in codified nonverbal behavior (Ekman and Friesen (1969), Mantovani (1996), Hiltz (1978)). This has been termed social feedback. Meanwhile, the speaker may use body language to emphasize her point, describe location, shapes, or time, mark the delivery of information, cite the interlocutor's contribution, seek a response, or coordinate turn taking (Bavelas et al. (1995), Hiltz (1978), Goldin-Meadow (1997)).

Section 1.4 Consequences for CSCW

1.4.1 Lean Communication Environments

People who regularly engage in Computer Supported Collaborative Work know that the communications media available to them are poor in content as compared to face to face communications (Hiltz (1978), Walther (1995), Picard (1995), Picard and Cosier (1997)). There is a well-documented drop in the effectiveness of communication when individuals interact through a computer interface. An extensive survey of the documentation concerning communication effectiveness in CSCW was performed by Walther (1995), who pointed out the immediacy of feedback and the number of channels used as determinant in fostering better communications in face to face exchanges than in computer-mediated exchanges. Picard (1995) also pointed out that users of lean communication media (such as CSCW tools) are prone to misunderstandings. Moreover, it seems that the users of CSCW tools are aware of the "leanness" of computer-mediated communications and tend to prefer "rich media" (i.e.: face to face dialogue) to

communicate in highly sensitive situations, or "highly equivocal information" (Walther, 1995).

It follows that, in order to restore the rich quality of face to face exchanges between CSCW users, several communication channels need to be added to the written text commonly supported by CSCW tools. In fact, users have already taken these matters into their own hands, and now commonly use emoticons in chat situations or emails, which was not the case in the earlier days of CSCW (Hiltz, 1978). Emoticons are intended both to provide immediate feedback to the writer (speaker) while letting her continue, and to elucidate the point made by the writer so that no misunderstanding can occur between the parties involved (Fridlund (1994), Picard (1995)). However, emoticons (at least English emoticons) can only express a limited number of very stylized expressions, and so should not be considered a final substitute for other types of media richness improvements (Picard, 1995).

Emoticons can only represent deliberate nonverbal behavior (see Chapter 2), and they cannot be coordinated with "speech" (or text, whichever method is used as the primary information exchange channel) as gestures are in face to face interactions (see Chapter 5). Therefore, emoticons have an inherently limited communicative value, and they cannot provide an end-all solution to the problem of media leanness in computer supported interactions. In order to provide a better, and more usable, communications channel for business environments, the NBCT should not rely only on emoticons, but

instead seek to establish a more sophisticated and realistic mode of information exchange.

An important requirement for realistic information exchange was introduced in Section 1.3. Specifically, the importance of gestures communications was outlined, and it was argued that gestures are an integral and necessary part of communications. There is therefore a clear incentive to seek to improve the transmission and representation of nonverbal behavior in CSCW. This will increase the number of communication channels available to users in CSCW situations, so that people should be able to interact more effectively in distributed environments. It is expected that this enhanced interaction, as explained in Section 1.2, will lead to the development of both trust and effective communication between CSCW users.

1.4.2 CSCW in Multinational Organizations: Communication Challenges

Multinational organizations face special challenges when it comes to communication between employees. It seems that these challenges are not currently addressed by collaboration tools, and one goal of the NBCT is to address them. In this section, the cultural challenges which face global companies are briefly considered (Section 1.4.2.1). The development of effective communications is presented as one potential solution to these problems (Section 1.4.2.2), and, lastly, further challenges pertaining specifically to the NBCT are considered (Section 1.4.2.3).

1.4.2.1 Globalization Challenges

The increased geographical distribution of organizations mentioned at the beginning of the chapter has been accompanied by a notable globalization of commerce and organizations. The development of multicultural firms over the last 15 to 20 years has put people from different cultures into close professional contact, both within and between organizations (Lewicki and Bunker (1996), Adams and Toomey, 1998). In order to take advantage of global markets, companies have realized that they needed to allow for the different cultures of their employees, suppliers, and clients to coexist and prosper along side each other. Books were published, covering the specifics of dealing with particular cultures, from the point of view of one's own (Watzlawick (1985), Carroll (1987)), both in professional and social contexts. For example, books for Americans travelling on business to Japan are available (e.g.: Gercik (1996)).

Multinational, and therefore multicultural, companies face problems which small, single country companies do not have to consider. Different laws, regulations, and taxation practices are only the first hurdle. Fostering collaboration and, more broadly, communication between culturally different people is the most taxing, demanding, and difficult part of successfully establishing durable and effective multicultural business relationships. However, it is also absolutely crucial if business is to be sustained in more than one country at a time (Belot (1999), Lagardère (1999), Mantovani (1996)).

In conclusion, respect for cultural differences, and acknowledgement of the specific communication and collaboration problems which they engender must be

addressed in order to ensure that successful cross cultural ventures be established. One way to address these challenges is to seek to foster effective communication within a multicultural organizations.

1.4.2.2 Effective Communication, a Path to the Solution

As was seen earlier in section 1.3, it is the development of the identification-based trust which governs the emergence of uniform goals between people. One of the challenges of managing large multinational companies is to ensure that people from different cultures have similar goals so that the business of the company and its objectives might be clearly and unequivocally understood by all (Belot, 1999). This can be very difficult, for example, in French society a company's responsibility is to the workers, who must be treated fairly, whereas in American society, the company's responsibility is to its shareholders, who must be enriched at all cost. This is clearly a stereotype, but it may reflect a reality which managers must deal with. Developing and fostering identification-based trust among co-workers can help bridge these differences and alleviate the misunderstandings which they can spawn.

However, as was seen before, identification-based trust, the highest level of trust, cannot be reached without significant interaction between individuals, and strangers can be limited in their interactive ability because they might not speak the same language equally well. Asynchronous CSCW tools can provide a safe environment in which people can take their time to compose messages in a foreign language (Adams and

Toomey 1998), but it was explained in Section 1.2.2 that providing real-time nonverbal behavior information will likely increase the communicative ability of people. Therefore, it is logical to strive to include social feedback, in the form of nonverbal behavior, as an aid in communication for cross-cultural computer-mediated communication.

Although providing nonverbal behavior information can enhance the communicative ability of people in CSCW environments, thus potentially leading to the development of the all important identification-based trust, it is not proven that the same effect can be achieved for CSCW tools working across cultural gaps. The reasons for this uncertainty are explored in the next section

1.4.2.3 A Combination of Challenges

Including nonverbal behavior information in CSCW tools poses a stiff challenge, especially if the information is to be shared by individuals from different cultures. This challenge results from the cultural dependence of the meaning of nonverbal behavior (Ekman and Friesen (1969), Labarre (1947)). These differences are offset somewhat by the universality of facial expressions which was revealed by Ekman (1982), but it remains that cultural display rules can significantly alter facial expressions, and that some gestures have very specific, learned meanings (Labarre (1947), Ekman and Friesen 1969), Birdwhistell (1970)). Consequently, there is much potential for misunderstanding foreign nonverbal behavior, and this topic will be further discussed in Chapter 2. However, it is

precisely to fill the need for better communication between foreign and distributed collaborators that the idea of the nonverbal behavior cultural translator arose.

The NBCT, used in its intended professional setting, aims to provide a tool for bridging the communication chasms which have been identified in this section. The NBCT will proceed in two ways. First by attempting to provide meaningful nonverbal behavior information as part of a real-time communication stream, and second by seeking to remove the cultural misunderstandings which arise in nonverbal communication. It is hoped that the enhanced communication experience provided by the NBCT will foster more effective communication between culturally different co-workers.

Section 1.5 Basic Requirements for the Nonverbal Behavior Cultural Translator

At this point, it is possible to establish some high level requirements for the nonverbal behavior cultural translator (NBCT). These requirements will serve as a guide for the rest of the content of this thesis. The NBCT will help fill some of the communication needs of large multinational firms, and so it will be used in business environments, more precisely to address the challenges which have been identified in this Chapter. The NBCT will not be useful by itself, and will have to be integrated into a larger CSCW tool as a complement to it. As such, it is not an independent product, and is not meant to be. There are two sets of requirements for the NBCT, one based on the expected needs of the users (user requirements), and one based on the functions which

much be filled by the NBCT in order to fulfill the needs of the user (functionality requirements).

1.5.1 User Requirements

The users of CSCW tools must be able to use the NBCT in an intuitive manner. Importantly, a user should be able to take advantage of the NBCT without needing to interrupt his or her normal activity (Vilhjálmsson and Cassel, 1998): ideally, the NBCT would work even without the user's awareness, much like a mail program which periodically fetches mail from a server. The NBCT must also be integrated seamlessly into the existing infrastructure of the tools in use. More importantly, the NBCT must make use of non-intrusive recording or displaying methods when recording is necessary. Lastly, the NBCT must provide useable representations of the translated nonverbal behavior to the users. The first requirement above (NBCT must be intuitive to use) and the last two requirements (non-intrusive recording and appropriate representation) warrant a brief discussion. Users impose no specific requirements on how the steps between the recording and the displaying are handled. These are technical requirements which will be discussed in Section 1.5.2.

When individuals engage in nonverbal behavior, this behavior does not interfere with their ability to communicate using the vocal channel. Indeed, as explained in Section 1.3.2, it is often in order not to interrupt the verbal channel that nonverbal behavior is used to exchange information. The NBCT must recreate this state of

communication in order to provide both a realistic and easily usable interface (Vilhjálmsson and Cassel, 1998).

It is important that the NBCT's users be able to leverage its functionality with limited reliance on wearable hardware, or on sensors applied to the body. This is in part due to the NBCT's intended use in business environments. There should not be a notable delay or a deliberate effort required in order to make oneself available to the NBCT for translation. Further, wearing hefty sensors and subjecting oneself to the contact of a recording device could make users unwilling to use the NBCT because of these sensors might not be comfortable to wear. Lastly, users wearing sensors might be self conscious of being monitored and might therefore not express themselves in a normal manner (Birdwhistell, 1970). This is why the NBCT must be able to perform all necessary recording of human behavior through a non-intrusive, preferably unnoticed manner. The preferred method will be to use a video camera, but it is not clear at this point how successful, accurate –and therefore useful– such recording will be for capturing all these different types of information.

Lastly, in order for the information gathered and translated by the NBCT to be used effectively by the users, an intuitive method of presenting that information must be devised. The NBCT cannot require that the user acquire a significantly new set of skills in order to be able to benefit from the NBCT. Rather, the user of the NBCT must be presented with information whose encoding is as natural as possible, in a way that is quickly understood by all. This will be discussed in Chapter 5. The information

representation must also be accessible in a manner which is similar to the way information in the nonverbal channel is accessible in face to face communication. In particular, the need of synchronization of speech and nonverbal behavior in order to provide real time, usable feedback must be addressed.

1.5.2 Functionality Requirements

This section will briefly describe the steps which are required of the NBCT in order to achieve its goal of translating nonverbal behavior between cultures. The series of steps is illustrated in Figure 1. A discussion of the steps follows.

The first step, as described both in the user requirements and in Figure 1, is to capture a representation of the nonverbal behavior of the users. This was briefly discussed above from the point of view of the user (Section 1.5.1), and will be discussed in further detail in Chapter 3. Let it suffice to say for now that the NBCT will require hardware capable of recording body motion and software capable of analyzing and characterizing this motion.

The second step is to interpret the recorded signal. Without this step, translation cannot take place. There are several options for how to perform this step, including the use of simple dictionary-like mappings between actions and meanings. What is important here is that it is the meaning intended by the performer of the nonverbal act,

rather than the meaning understood by the receiver, which will of importance in this endeavor. The reasons for this necessary distinction will be explained in Chapter 2 and 4.

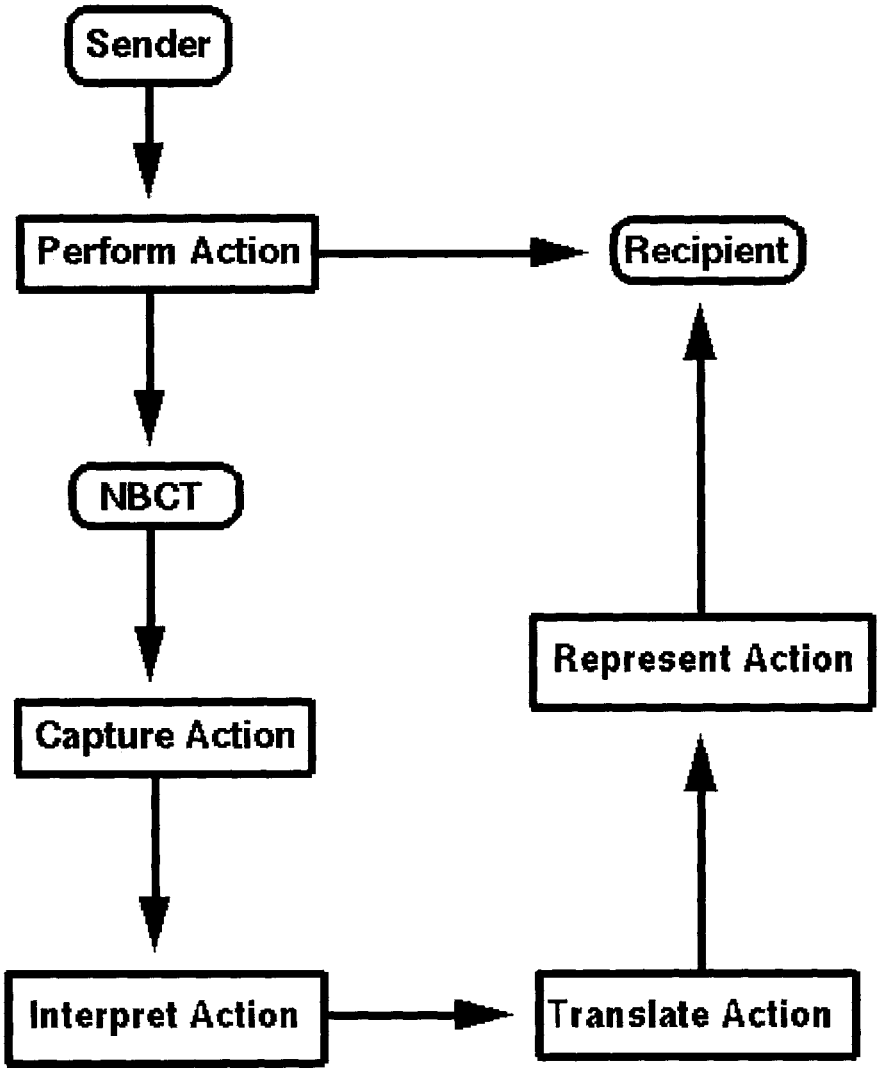

Figure 1. The path of information between the different parts of the NBCT. The arrow from "Perform Action" to "Recipient" shows the path of information if no interface is present (i.e.: face to face communication)

Next, the NBCT should be ready to translate the interpretation into a nonverbal behavior appropriate for the recipient. This nonverbal act is ideally based on the meaning extracted in the previous step, and will be chosen in order to convey as accurate a meaning as feasible (methods involving user feedback may be considered). This is dependent on the culture of the recipient, as will be explained in Chapter 2. This is precisely what is done in dictionaries, where the meaning of a word must first be conceptualized before it is translated into another word which has the same meaning in a different language. This will be further discussed in Chapter 5.

Once the representation has been chosen from a list of possibilities, it becomes necessary to display this representation. Just how the representation should be made available to the user is not clear, but there are several options which will be further discussed in Chapter 5. It is, however, clear that concerns about the type, synchronization with speech, and frequency of these representations need to be addressed explicitly. Ideally, the NBCT would be able to provide culturally translated representations of nonverbal behavior at a pace and level of subtlety and detail which would match face to face communication.

Section 1.6 Conclusion:

After introducing the concept of multi channel communication, and discussing the failings of CSCW tools at providing this type of communications, this chapter identified the need for better communication between employees of multicultural companies.

Harmonizing the goals of their employees is one of the stiffest challenges facing these companies (Section 1.4.2.1). This chapter described how the establishment of common goal is dependent on the emergence of a type of trust called identification-based trust. This type of trust is in turn dependent on the establishment of effective communications between the employees of a given organization. Section 1.4.2.2 suggested that effective communication could serve as a path out of this problems for international companies. The chapter concluded by providing some requirements for the NBCT. These requirements are important because they should guide the design of the NBCT, making sure that it remains usable by largely inexperienced business professional users. More importantly, the requirements sections highlighted the necessary steps to be taken in creating an NBCT.

As was seen in Section 1.5.2, the NBCT requires the implementation of several critical elements (capturing, interpreting, translating, and representing user movement). Each element depends on the earlier ones in order to function, and might require completely different sets of skills to implement. In fact, as will become evident throughout the next chapters, building the NBCT will require the collaboration of experts from several fields, including psychology, computer-based pattern recognition, and image rendition to name a few. These are potential difficulties if the NBCT is ever to be built, but one that is beyond the scope of this thesis. Systems which are dependent on all of their components to function (series systems) are inherently weaker than parallel systems. This constitutes a liability in the original phases of the conception of the NBCT (i.e.:

before all of the required technologies become viable), but it should not present a problem once the hurdles identified in this thesis are overcome.

Before embarking on the explorations of the theory and technology required to implement the critical steps listed above, it is necessary to explore culture and cultural differences, and to provide a concise and synthetic definition of culture. Whether such a definition can be provided is not clear, but an attempt is made in Chapter 2. This is necessary to be able to be able to find an integrated solution to the problem of translating nonverbal behavior between cultures.

Chapter 2 Culture and the Meaning of Expressions

"It is not possible for anyone to see anything of the things of that actually exist unless he becomes like them"

The Gospel of Philip

In order to be able to understand the specifics of translating nonverbal behavior between cultures, it is important to use a potent and highly synthetic definition of culture. This chapter will attempt to provide such a definition in Section 2.1. The goal is to provide an efficient manner of considering culture, rather than to devise yet another cumbersome definition. The goal is not to provide a definitive or even ethnographically satisfactory definition of culture. The author feels that the present definition provides a convenient framework to express the needs and requirements of the NBCT. In particular, the dependence of the meaning of events on culture will be discussed in Section 2.2.

As explained in the previous chapter, nonverbal behavior is an important part of communication, and so the dependence of nonverbal behavior on culture will be explored in Section 2.3, following a brief theoretical overview of nonverbal behavior itself. Context, and its effect on meaning will also be considered in Section 2.3.2 because it

adds a level of complexity to the task of the NBCT. Finally, there will be a discussion of the crucial concept of display rules in Section 2.4.

Section 2.1 Culture as a Frame of Reference

Authors have defined culture in several different ways, most often in order to suit their specific needs. Specialized definitions allow for specific analyzes, emphasizing a particular aspect of human behavior, or addressing the particular concerns of a class of scientists. In general, definitions are formulated with metaphors or technical jargon, or a combination of both. A survey of some theoretical definitions is provided by Mantovani (1996). Ethnographical (descriptive) definitions are available from Watzlawick (1987) and Carroll (1987).

For the purposes of this thesis, and in the interest of finding a convenient language to describe the challenges facing the NBCT, culture will be defined in mathematical terms. Such a definition will allow for the use of both metaphors and mathematical ideas to describe cultural differences, interpersonal interactions, cultural translations, and cross-cultural education (Section 2.1.1.3).

2.1.1 Definition of Culture

This section will provide a succinct definition of culture (Section 2.1.1.1), introducing a new model for culture. This definition will be supplemented by examples

(Section 2.1.1.2). Finally, some of the perceived benefits of this new definition will be presented in Section 2.1.1.3.

2.1.1.1 The Definition

At any given time, for a given individual, the culture of that individual can be represented as a vector basis. Each vector in that basis corresponds to a fundamental dimension of the culture, quite in the same manner as in a physical frame of reference. Cultural bases are n-dimensional in general, and they are made up of the union of n different cultural vectors of length m (C_A is the cultural space for a person A):

$$C_A = C_1 \cup C_2 \cup C_3 \cup C_4 \cup \ldots C_n$$

Events (communicative acts) are not considered to exist within a culture, rather they exist in an event space. The event space is unique and is made up of all possible cultural vectors, and so it has dimension k, where k is the total number of cultural vectors, and is assumed to be finite. This is an important assumption which is justified because there are a finite number of individuals on Earth at any time, and therefore a finite number of possible cultures.

Events acquire a meaning within a culture (or coordinates within that culture) when they are projected onto that culture's basis. Thus the matrices which allow for transformation between the cultures and the event space are defined with respect to a canonical space: the event space. Mathematically, this is not strictly necessary, but it is convenient to go through the event space for the purposes of the NBCT (as will be

explained in Chapter 4). In this definition, events can be any communicative act, even if it is not intentional. A further discussion of communications and of what constitutes a communicative act can be found in Sections 1.2, 1.3, 1.4 in the previous chapter, and in Section 2.3 in this chapter.

This definition should not be taken in a strict mathematical sense, but rather as a metaphor which allows concise and practical discussions of cross-cultural exchange. The object of this definition is to use terms and concepts likely familiar to those potentially developing the NBCT, which is not the case for the definitions reviewed by Mantovani

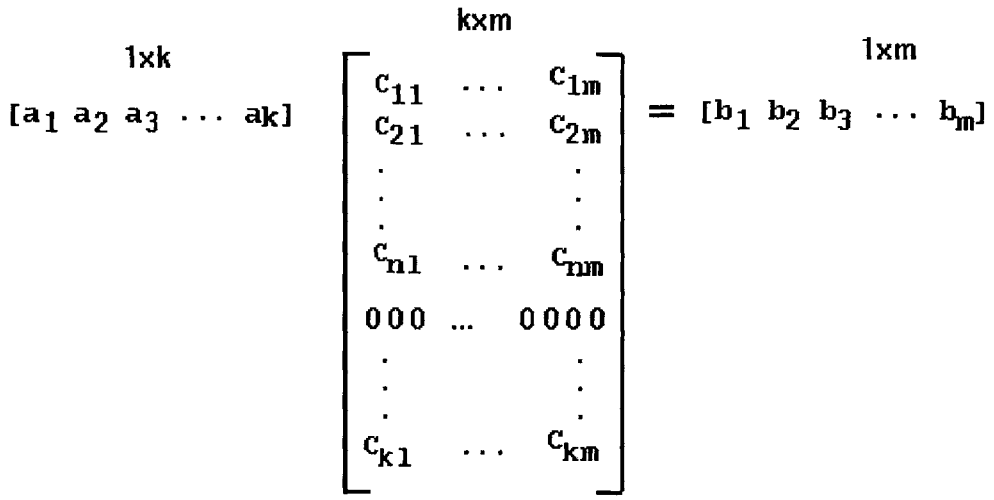

Figure 2. Matrix multiplication showing a change of coordinates from event space to cultural space. a_1 to a_k are event space coordinates and b_1 to b_k are cultural space coordinates

(1996). A "matrix multiplication" shows how event space coordinates are changed to cultural coordinates (Figure 2). It is important to note that since the cultural matrix has dimension (n x m), additional, trivial (zero) vectors will be needed to make it (k x m).

This is how some information about an event can may not be accessible to another culture (i.e.: the corresponding row in the culture matrix is zero).

As compared with other available definitions of culture (Mantovani, 1996), this definition has the important advantage of being very succinct and internally consistent. However, the definition should not be taken literally, as it is very unlikely that a numerical decomposition of cultures can (or should) be produced. In order to illustrate the definition, and also to demonstrate its usefulness, an analogy with physical frames of reference is drawn.

2.1.1.2 An Analogy and Some Examples

In the common human physical frame of reference, there are three dimensions: one vertical (up-down), and two horizontal (left-right and front-back). These dimensions provide a way to describe the physical world, e.g.: objects are close, far, high, low, right, left, or any combination of these. In a cultural frame of reference, the dimensions correspond to descriptions of events (this is further explained in Section 2.1.2), rather than to physical descriptions or positions of objects. For example, as illustrated in Figure 3, typical cultural dimensions might include nervous-relaxed, angry-pleased, uneasy-confident, or worried-carefree.

Each of these axes is a pair of antonym adjectives which might describe one aspect of an event. Figure 4 shows a three dimensional cultural subspace with an event.

It is very important to realize that a particular culture's nervous-relaxed axis might be orthogonal to another culture's. This last point will be further discussed in Section 2.1.2, and it indicates the limits of the analogy with physical space.

$$\begin{bmatrix} c_{11} & \ldots & c_{1m} \\ c_{21} & \ldots & c_{2m} \\ \text{nervous} & \ldots & \text{relaxed} \\ \text{angry} & \ldots & \text{pleased} \\ \text{uneasy} & \ldots & \text{confident} \\ \text{worried} & \ldots & \text{carefree} \\ \vdots & & \vdots \\ c_{k1} & \ldots & c_{km} \end{bmatrix}$$

Figure 3. Example of a typical cultural matrix

Pursuing the physical frame analogy, it is possible to say that humans use their physical frame of reference to evaluate their positions with respect to objects in the physical world. Once the object is perceived by any combination of the senses, its location can be determined by the brain. In other words, the brain assigns coordinates to that object, starting from an origin which is usually located somewhere in the body. For example, our brain can let us know where a door handle is with respect to our eyes: down, front, and left. However, it can also be more precise: 2 feet down, 1 foot in front, and 1 foot to the left. This last description is analogous to assigning coordinates to one point of the door handle, i.e.: (2,1,1), the origin being a point in our body chosen by the brain. Our brain can also tell us how big an object is and how fast it is moving with

respect to ourselves. Individuals do this naturally and intuitively, having learned to evaluate distance as a practical survival skill. However, not all people possess this ability. In particular, astigmatism is a vision ailment which impairs the perception of distances. In other words, astigmatic individuals are unable to assign accurate coordinates to the objects which surround them.

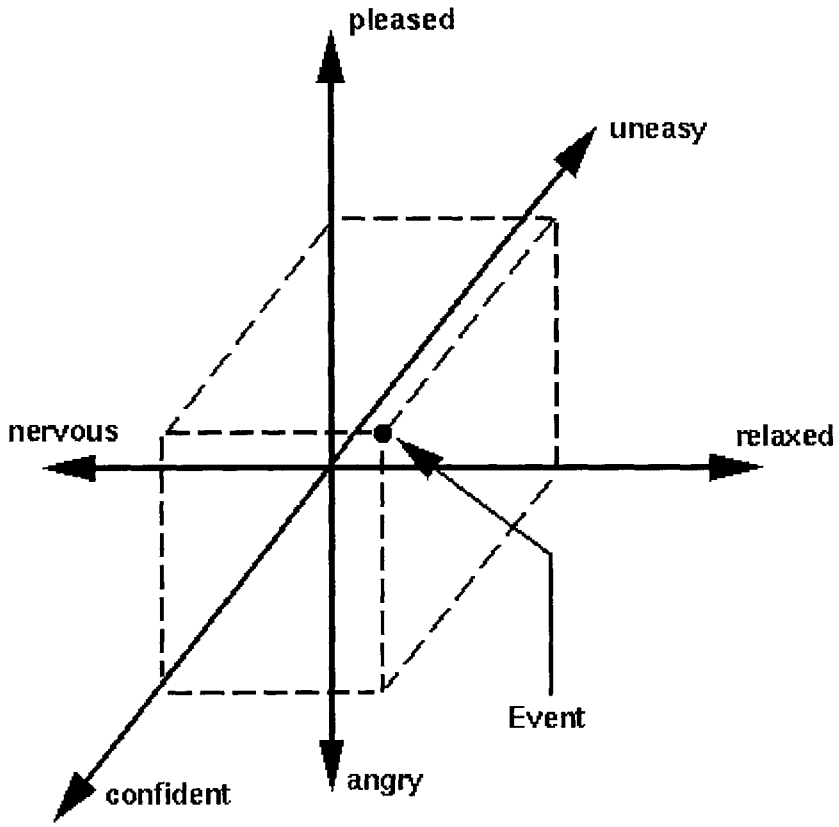

Figure 4. A view of a cultural sub-space for a given culture. The event shown has been assigned coordinates in that sub-basis.

In the cultural model described in Section 2.1.1.1, events are said to exist outside of the cultural frame of reference of a person: in the event space. This is intended to

mean that once an event is performed, it is like a soap bubble. A soap bubble is produced by a person, but once it leaves that person, it has no bounds to its creator. Its coordinates can be determined with respect to an arbitrary frame of reference (e.g.: "bubble space") which might have as its origin at the center of the bubble. This space is related to the physical frame of reference of the people in attendance through simple matrix multiplication. All in attendance can see the bubble, and so evaluate its position, size, and speed (its physical characteristics) according to their individual physical frame of reference. In general, it is possible that some disagreement might occur over its characteristics if no objective evaluation method or tool is used.

Similarly, when an event occurs, all can witness it and evaluate it according to their individual cultural frame of reference, performing "matrix multiplication" of the event space coordinates (Figure 2). This evaluation takes place as each individual assigns "coordinates" to that event along the cultural axes. The matrix multiplication used for coordinate assignment is filled with cultural vectors (Section 2.1.1.1). Those cultural vectors are really the embodiment of what Ekman and Friesen (1969) called "display rules". Display rules are described in Section 2.4. Once again, it is important to understand that this is not strictly speaking "matrix multiplication" since it is highly unlikely that such a synthetic representations of cultures could be produced. However, this is very powerful analogy which will allow for a convenient way to formulate the problem of the NBCT. At this point, an example is necessary to clarify this analogy.

Let us consider two individuals shaking hands while smiling; this constitutes an events. The physical nature of the event is clear and unmistakable: it has an objective reality. What might not be, however, is the meaning that each of the parties attaches to that handshake. One might use his display rules to project the "handshake-with-smile" event onto his salutation-ignoring, his friendly-unfriendly, his joking-serious, and his casual-formal axes. He might have other axes at his disposal, but his display rules tell him to decompose the action only with along these vectors. He then determines that the event has salutational, friendly, formal, and serious components, but no nervous-relaxed coordinate. The other might use her own display rules and project the event onto her angry-pleased, bored-interested, and uneasy-confident axes, thus finding the event to have angry, bored, and confident coordinates, but no happy-sad or salutation-ignoring coordinate. The two people involved will therefore have different understandings of the event. This example is clearly imperfect, however, it is merely intended to provide insight into a much more complicated situation.

It is important to recall that once it is performed, an event only has coordinates in the event space. Events only acquire cultural coordinates once an individual places one in his or her frame of reference. This is because events exist only when they have an originator (and, ideally but not necessarily, a receiver). This is an important difference between events and objects, since the latter exist even in the absence of perception by anyone. This analogy has run its course, and has amply illustrated the definition. It is now appropriate to consider some of the benefits gained from this definition.

2.1.1.3 Some Benefits of the Present Definition

The definition of culture as a vector basis allows for the succinct discussion of usually spiny problems (Mantovani, 1996). For instance, cultural differences can be represented by how linearly independent two cultural frames are. Even dimensions (or vectors) which have the same name (e.g.: nervous-relaxed) need not be parallel. This allows for the same event to project differently onto different frames of reference, not only in terms of coordinates along a vector, but also in terms of the dimensions themselves. For example, consider an event which is nerve-wracking to a Japanese (high coordinates on the nervous side of the Japanese nervous-relaxed axis). This same event might be relaxing to an American (project onto a different part of the American's nervous-relaxed vector, i.e.: the vectors are not orthogonal). This event might be cause for joy to a Papuan, and not cause for any nervousness or relaxation (i.e.: the nervous-relaxed axis of the Papuan is orthogonal to the two others'). Expanding this idea over all of the dimensions of a cultural basis, one can see how cultural difference is akin to linear independence of cultural vectors.

Still abiding by the culture model, interpersonal interactions are represented in the following manner: one party outputs an event using his or her frame of reference to "encode" it (this is further explained in Section 2.4); that event is then evaluated by the other parties involved in the interaction, according to each party's frame of reference. In mathematical terms, considering two communicating individuals A and B, with B the sender of an event and A the receiver, this statement can be written as follows (as inspired by Figure 2):

$$E_B * [T_{BE}] = E_E \qquad Eq.\ I$$
$$E_E * [T_{EA}]^{-1} = E_A \qquad Eq.\ II$$

where E_B is the event output by B evaluated in B's frame of reference, E_E is the event in the event space, and E_A is the event evaluated in A's frame of reference. T_{BE} and T_{EA} are maps to and from the event space. T_{EA} is a cultural matrix for individual A's culture, as shown in Figure 2, and T_{BE} is the inverse of T_{BE}, the cultural matrix for B.

Further, cross-cultural translation can be modeled as the projection of an event onto the sender's frame of reference in order to discover the intended meaning, followed by the projection of the event corresponding to that meaning onto the receiver's frame of reference. These two steps are described in greater detail (and mathematically) in Chapters 4 and 5 respectively. Finally, learning a culture can be reduced to simply acquiring new basis vectors, or, in other words, increasing the dimension of a person's cultural basis (see Section 2.1.2.1 for further details). The definition of culture proposed in this chapter is satisfactory and complete because it describes cross cultural exchanges and differences, as well as the evaluation of events by individuals (Shweder and Sullivan (1993), Mantovani (1996)).

2.1.2 Consequences of the Definition

The definition of culture provided in Section 2.1.1 has several advantages: a) the individual is the atomic element of culture, b) learning about another culture is simply conceptualized as adding dimensions to a basis (a) and b) are described in Section

2.1.2.1), c) events (e.g.: communicative acts) and their meanings are separated from cultures (Section 2.1.2.2), d) interpretation of events and of their meanings can be viewed as mathematical projections onto a person's cultural basis (Section 2.1.2.3), and e) events can be decomposed along the cultural dimensions described in Section 2.1.1 (Section 2.1.2.4).

2.1.2.1 Multi-cultural Individuals

It is important that the individual be considered the atomic element of a culture because the traditional concept of culture as linked to nationality is quickly fading (Carroll, 1987). The increased contact between foreigners and the expatriation of individuals has created numerous people whose culture has been modified by their

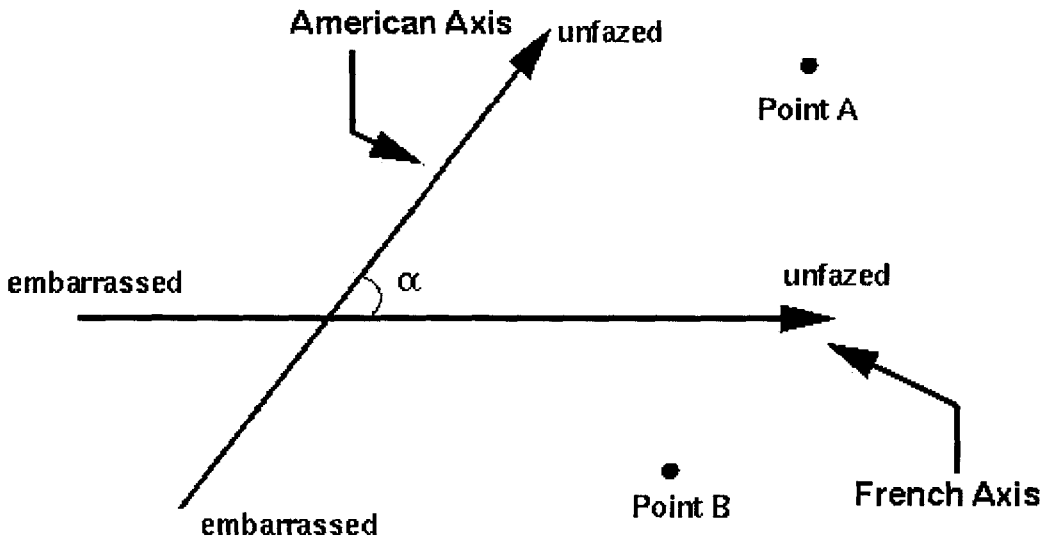

Figure 5. Schematic of the embarrassed-unfazed axes for the Frenchwoman. Points A and B represent events, and the angle α can have any value. A leaves both French and Americans unfazed, while B embarrasses Americans without bothering the French.

experiences in foreign countries (Carroll, 1987). These are the so-called multi-cultural individuals. Through contact with other cultures, they have acquired additional vectors: they have learned about that culture. In mathematical terms, a person A with culture C_A (C_1, C_2, C_3, ..., C_n) can gain some vectors from a culture C_B (C_{n+1}, C_{n+2}, C_{n+3}, ..., C_{n+k}) by simply adding them to his basis, creating a new culture, i.e.: C_A' (C_1, C_2, C_3, ..., C_n, C_{n+2}, C_{n+3}). In less abstract terms, after exposure to the American culture, a Frenchwoman might acquire the embarrassed-unfazed America vector in addition to the same French vector. As any two vectors, these vectors define a plane which allows her to understand both those events which are embarrassing to Americans and the different events which embarrass the French (Figure 5). In this model, multi-cultural people have their own culture, which is unique according to the uniqueness of their vector basis. This does not mean that several, or even many, people cannot share the same cultures. However, it does allow for the existence of cultures which are only shared by one or two individuals. This can occur if the individual experiences and particular education of these individuals are sufficiently unique.

2.1.2.2 Events

The separation of events from cultures is essential for the development of the nonverbal behavior cultural translator, as will become evident in Chapters 4 and 5, because it allows for simple mappings between meanings and events. Briefly, once the cultural "coordinates" of an event have been found in the sender's frame of reference, then an event with the same coordinates can be designed in the frame of reference of the

receiver. If events exist outside of a particular culture, then they are theoretically accessible (or visible, however imperfectly) to all cultures. In other words, some events may project without loss of information onto a particular basis, and yet have no projection onto another cultural basis. Models which include the events within the cultural framework (Sahlins (1985), Beckmans (1996), Labarre (1947), Montovani (1996)) effectively block access to these events from people outside of that culture: only those which are part of the culture can see the event because only they possess the cultural "vocabulary" to describe it. These models therefore introduce redundancy as similar events need to be described in different cultures. This separation represents a significant improvement of the present model over previous culture models, at least from the point of view of building an NBCT.

2.1.2.3 The Meaning of Events

In this model, the meaning of an event is simply the projection (in a mathematical sense) of the event onto an individual's cultural vector basis. The meaning is thus determined at the individual's level, rather than being an inherent characteristic of the event itself. The analogy with physical space provided in section 2.1.1 gave an example. What the individual can "see" from that event is determined by the extent to which her basis can describe the particular event, or, in mathematical terms, by how many trivial vectors must be added to her basis to perform the matrix multiplication described in Figure 2. Much like the description of a four dimensional object is difficult for humans who live in three dimensions, so a person's perspective is limited by her cultural frame of

reference. Moreover, given extra dimensions, as is the case with multi-cultural individuals, a human being can inspect a particular event from different points of view. This was the case of the Frenchwoman mentioned in Section 2.1.1. She was able to evaluate situations potentially embarrassing for Americans, something which an insular Frenchman would not have been able to do.

2.1.2.4 Decomposition of Events

Finally, it is also important to understand that individual events can be decomposed into non-redundant atomic cultural coordinates as they are projected onto individual frames of reference. This means that this method gives much flexibility for interpreting and decomposing events within a cultural frame of reference (see the handshake-with-smile example in Section 2.1.1). This is unlike other models which require the consideration of events as a whole, without setting either a limit or a method for decomposition into elemental units (Mantovani (1996), Beckmans (1996), Labarre (1947)).

Traditional culture models are cumbersome, and they do not lend themselves to the divide and conquer method which the author favors for implementing the NBCT. It was therefore necessary to provide a more usable cultural model. The specific exercise of determining and understanding the characteristics of this new model provides a solid theoretical groundwork for thinking about NBCT implementation. While the model is by no means complete, its abstract nature is consistent with the needs and assumed preferences of those who might one day implement the NBCT. Section 2.1.3 explores

the specific consequences of this new culture model for the NBCT, and provides a succinct rewording of the problem of translating nonverbal behavior.

2.1.3 Consequences for the NBCT

Clearly, some cultures share many basis vectors, while others are entirely orthogonal to each other. The very large number of combinations of parallel and orthogonal bases as well as individual vectors is meant to span all possible cultural subgroups, thus allowing for subtle differences between groups and individuals. Loosely speaking, if a computer could be programmed with to decompose events using the cultural vectors of a particular cultural group, it can associate a meaning to those events using "display rules' (see Section 2.4). That meaning could then be translated. This is a mere rewording (albeit a powerfully simple one) of the NBCT steps outlined in the first chapter (Figure 1) in terms of the present cultural model.

Of course, this is very difficult to achieve in a way that will look credible because it is akin to teaching a computer to use "every day knowledge", also known as common sense. As described by Dreyfus (1992) Doug Lenat of Cycorp has identified that teaching computers how to use "every day knowledge" is extremely difficult, due to the current lack of understanding concerning:

1. How everyday knowledge must be organized so that one can make inferences from it,
2. How skills or know-how can be represented as knowing-that, and
3. How relevant knowledge can be brought to bear in particular situations (McCarthy, 1996)

However, despite this serious challenge, it is important to realize the potential of the new culture model presented in this chapter, and to attempt to leverage this power to precisely outline the problem so that one might attempt to solve it. "First order" attempts at solutions will be provided in the next chapters, as was explained in section 1.5.2.

Despite the obvious difficulty of the task at hand, it is important to remember that learning about another culture, whether by a human or a computer program, can be simplified to adding cultural vectors (Section 2.1.2.1). Further, translating is represented by successive projections to different frames of reference (Section 2.1.1). These very simple and powerful metaphors are provided by the current culture model for the purpose of building a NBCT, allowing for a concise re-formulation of the problem at hand.

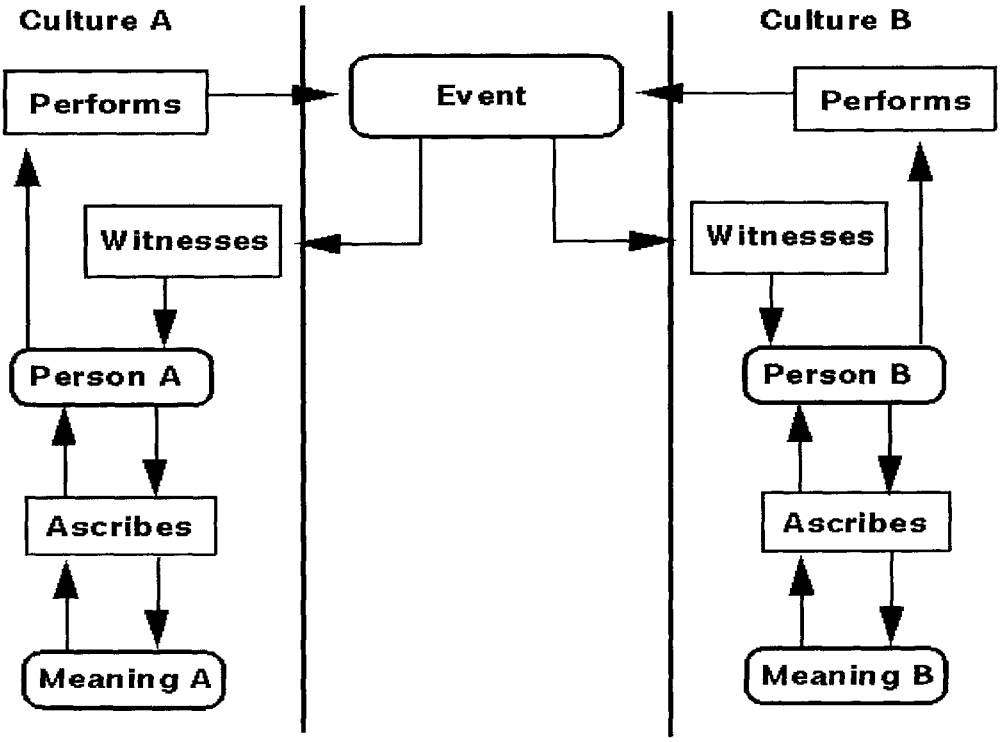

Figure 6. A representation of the relationship between culture, events, and meanings.

Section 2.2 Cultural dependence of meaning

This topic was briefly touched in Section 2.1.2.3, but a further discussion is warranted. From the definition of culture given in this chapter, a general schematic illustration of the relationship between individuals and the meaning of events can be produced (Figure 6). The goal of this figure is to highlight the fact that the interpretation of an event takes place at the level of the person. This interpretation occurs within, and subject to, the restrictions imposed by his or her cultural frame of reference. The assignment of meaning is therefore an intrinsically personal activity, as was described by the cultural model developed in Section 2.1.1.1. Furthermore, the figure points out that there are in general several possible meanings for the same event, one for each person/culture involved in the exchange. In general, meaning A and meaning B need not be at all related to each other. This is a direct consequence of the culture model outlined in Section 2.1.1.

Looking at Figure 2, it is easy to see that it is not until there are common vectors (or common dimensions) in A's and B's two bases that a common meaning can be reached. Only then will the projection of the event onto both basis will be equivalent (i.e.: the events will have equal linear decompositions along equal vectors in the cultural bases). This is very similar to the problem that was described by Mantovani (Mantovani, 1996) with respect to the need for reciprocity in communication in general (Section 1.2). Mantovani claimed that effective communication could only occur once a common vocabulary, or "shared symbolic order", was established. To say that two people need to

share the same friendly-aggressive axis is only a succinct manner of saying that they must reach an understanding about what constitutes friendly or aggressive behavior.

In simple terms, A and B need to reach a common ground for the evaluation of an event until they can both have effective communication about this event. They need to learn about each other's culture (and acquire some of the other's cultural vectors) in order to understand an event as the other does. After this learning takes place, their cultural basis is changed. They then have the option of evaluating events according to either vector or according to both. This was the case of the Frenchwoman in Section 2.1.2.1. The effective communication challenge described in Chapter 1 has gotten more difficult with the inclusion of culture as an added hurdle and variable.

There are numerous stories of examples of the misunderstandings which can result from the projection of an event onto one's own cultural basis. Indeed, whole books have been written on this topic (Gercik (1996), Watzlawick (1985), Carroll (1987)). A specific, stereotypical example can be found at the end of this chapter, in Section 2.4. In general, it can even be that an event which in one culture is cause for joy or pride would elicit the opposite emotion in another culture. More commonly however, an individual simply does not know how to evaluate a particular situation, or what the other person is trying to convey. This is often the case for gestures and nonverbal behavior.

Section 2.3 Cultural Dependence of Nonverbal Behavior

The risks for misunderstandings between cultures extend to nonverbal behavior. This is because a significant portion of nonverbal behavior is learned and is used as language (Ekman and Friesen (1969), Labarre (1947), Birdwhistell (1970)). Further, once nonverbal behavior is performed by a human being, it becomes an event. This event will then be interpreted by the interlocutor as described in Section 2.1, according to a projection onto the interlocutor's frame of reference. This process occurs irrespectively of how linearly independent the sender's and interlocutor's frames of reference might be. Meaning evaluation always occurs at the level of the individual in this model.

There is therefore a double dependence of nonverbal behavior on culture, first in the meaning that the sender intends or understands in performing an act (not all acts are deliberate, as explained in Section 2.3.1), and second in the understanding which the receiver extracts from that act. A brief introduction of nonverbal behavior is provided in the next section (Section 2.3.1), emphasizing cultural differences in gestures. Section 2.3.2 contains a discussion of the dependence of meaning of nonverbal behavior on context.

2.3.1 Nonverbal behavior, a theoretical overview

In order to discuss the specific challenges which affect nonverbal communication, it is relevant to delve a little further into the specifics of nonverbal behavior. The seminal paper on nonverbal behavior was written by Ekman and Friesen in 1969 (Ekman and

Friesen, 1969). His classifications have changed little since then, even though they have been refined by further studies. Because they are both accurate and conveniently simple, these classifications will be used in this thesis.

The first important division in nonverbal behavior is between deliberate and non-deliberate behavior. Deliberate behavior is behavior through which one agent intends to express or convey information. Non-deliberate behavior refers to the agent's behavior unintentionally expressing something about the agent. In both cases, the meaning must be inferred by the interlocutor (Beckmans, 1996), i.e.: the meaning is not intrinsic to the event in either case.

A further classification was proposed by Ekman and Friesen (1969). They classified nonverbal behavior into five main categories: emblems, illustrators, regulators, affect displays, and adaptors. They also defined the ideas of "usage" –the circumstances of use, "coding" –the rules which explain how the behavior contains or conveys information, and "origin" –how nonverbal behavior evolved– of nonverbal behavior. Of these three concepts, only the usage and coding of nonverbal behavior are of particular interest for this study. These will tell how a particular nonverbal act is used and how a meaning is encoded in it. This is crucial to the NBCT, since it must determine the meaning of a nonverbal act before being able to translate it. The "origin" of nonverbal behavior is largely irrelevant to this study except in determining whether a particular behavior is learned or innate. If a behavior is learned, it will likely be culturally specific, and if it is innate, it might be universal in coding (Ekman and Friesen, 1969). This will

affect whether not an act needs to be translated. While a detailed discussion of the characteristics of each type of nonverbal behavior is not relevant for the purposes of this thesis, a brief overview of the characteristics of emblems (Section 2.3.1.1), illustrators (Section 2.3.1.2), and affect displays (Section 2.3.1.3) is required. These three types of nonverbal behavior are of particular interest because their origin is culturally specific.

2.3.1.1 Emblems

Emblems are deliberate "nonverbal acts which have a direct verbal translation, a dictionary definition", and this definition is "well-known" by the members of a cultural group (Ekman and Friesen, 1969). Emblems are so precise that they can substitute for language, if needed. They are nearly always intentional. For example, the "boring" gesture in French, which is performed by rubbing one's knuckles against one's cheek, is an emblem. This is not a natural gesture which occurs by chance, rather it actually constitutes a subtle and deliberate play on words codified into a gesture. Emblems are the most culturally specific type of nonverbal behavior because they are learned behavior. They are also the easiest to translate since they are deliberate, and have a clear, one to one correspondence (or a single projection) with a meaning in a given culture: they are the most unambiguous of all types of nonverbal behavior.

2.3.1.2 Illustrators

Illustrators, as their name suggests, are used to illustrate speech, and are directly tied to what is being said (Ekman and Friesen, 1969). They are used to emphasize, sketch a thought, point, and show a spatial relationship or a movement (Efron (1941), Bavelas et al. (1995)). They are meant to add content to the spoken information, and cannot act as a substitute for it. For example, a common illustrator is to point at the object being discussed, or, as an illustration, their equivalent in the voice channel are intonation, loudness, or inflection. The type of illustrator used is variable with culture, and some cultures use more illustrators than others (Ekman and Friesen, 1969), also their coordination with speech is variable (Streek, 1993). These gestures, while intended to communicate, are not always deliberate. People use them out of habit, and it may be that some illustrators in one culture could be emblems in another (or vice-versa), leading to potentially embarrassing or damaging misunderstandings.

2.3.1.3 Affect Displays

Affect displays are mostly produced by the face, and are usually non-deliberate expressions of an inner emotional state. Much has been written about affect displays and their importance to computing, especially by affective computing specialists (Picard (1997) , Picard (1995)). One of Ekman's most significant contributions to the study of nonverbal behavior was the identification of universal, non-deliberate facial behavior (Ekman, 1982). However, it is important to realize that cultural display rules can modify, dissimulate, or reverse, the reflex-based affect displays. In this sense, affect displays are

both deliberate and culture specific; for example some cultures will smile to hide fear or embarrassment. A recent breakthrough in affect recognition has made it possible to distinguish between "real" and "fake" smiles (or facial expressions in general) (Bartlett et al., 1999). This particular technological and psychological breakthrough will be discussed in Chapter 3.

2.3.1.4 Consequences for the NCBT

Section 2.3.1 has explained up until now how nonverbal behavior is classified, and how it is culturally variable. Nonverbal behavior, as explained in Chapter 1, is an integral part of effective communication. Since the goal of the NBCT is to help foster effective communication, it will need to provide a manner for limiting the misunderstandings which can arise from the cultural variability of nonverbal behavior.

The review of the classifications of nonverbal behavior offered in this section allows future designers of the NBCT to have a vocabulary for dealing with nonverbal acts. The realization that there are at least three types of nonverbal acts should allow future NBCT designers to focus on specific types of acts as their skill progresses. For example, it is conceivable that a first version of the NBCT would contain only a dictionary for emblems. Later versions will ideally be able to recognize all three types of nonverbal behavior, and those acts in particular which are used during business interactions, in order to provide accurate and useful translations of nonverbal behavior between cultures.

Of course, nonverbal behavior, much like spoken communication, is inscribed within a larger communicative context. This context affects several parameters, including the choice of words or acts, but also their very meaning and significance of an act. There are rules which are either implicit or explicit with every communicative situation. This warrants further investigation.

2.3.2 Dependence of meaning on context

In general, (whether justifiably or not) one does not address the President of the United States in the same way as a taxi driver. Moreover, one may address the same person differently depending on the situation (or context). Therefore, at least two context variables can be identified, both of which might modify both the verbal and nonverbal communicative behavior of an individual: the "identity" of the interlocutor, and the "situation". In this case, "identity" is loosely defined as "who the person is" (e.g.: the boss, a coworker, a janitor, the Pope). Characteristics of a situation can include location, time, and preceding events among others.

The effect of context on communication can be in the choice of words or acts, but more importantly for this study, it can be on the meaning of particular signals. In general, the "identity" of the interlocutor will affect the choice of signals, whereas the "situation" (or environment) will influence the meaning of a particular signal (along with, to a lesser extent, its selection). The question really is: how can one tell between tears of joy and tears of pain or sorrow? The answer lies in clearly identifying those events which

lead to the tears, as well as who the person crying is. This will determine the context of the crying. This context will in turn provide some clues as to which tears are being cried (joy or sorrow).

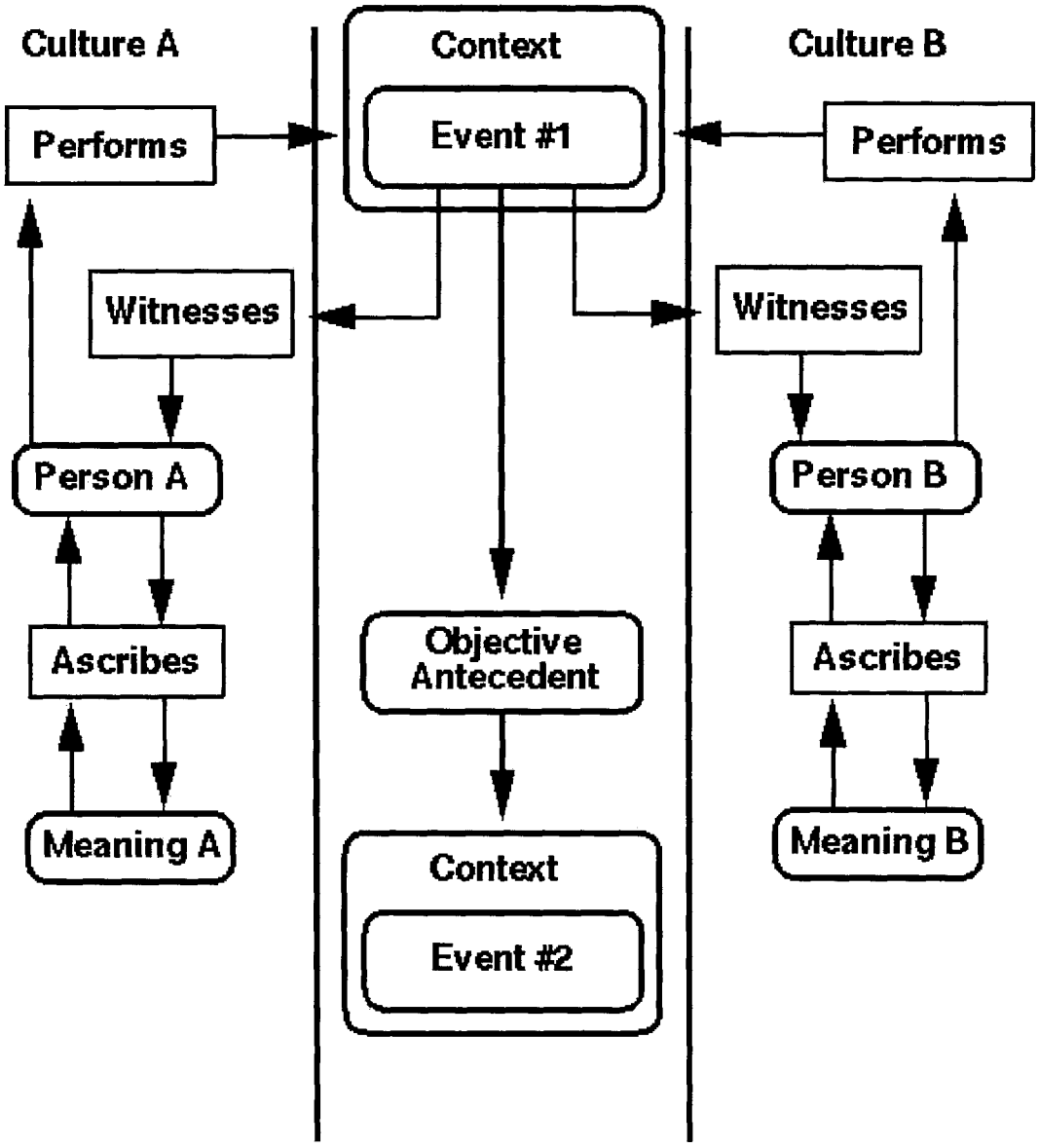

Figure 7. An illustration of Context, with respect to Objective Antecedents, Culture, and Events. Note that Event #1 has no context, since it is the first one of the interaction between A and B.

Communication signals are decoded by the receiver within a context. It is therefore natural to try to identify the characteristics of a situation that can yield information about that context. This is a very difficult problem for a computer, as it requires determining a method for information selection from a large influx of information (McCarthy, 1996). However, a potential simplification lies in that no evidence was found that context is or is not culturally specific per se, even though it can vary from one person to the next for the same event. This is commonly called "perspective". "Perspective" can force context to be defined at the level of the indiviudal, i.e.: each has his/her own perspective. Thus context, if defined loosely, will be a variable at the level of the individual. However, such refinement (defining context at the personal level) is not necessarily relevant for the purposes of the NBCT. In fact, it might be worthwhile and simpler to consider the objective antecedents to an event as its context, and to decode the event according to that context. In this case, event and context are associated as a non-separable pair. The objective antecedent to an event is defined here as the indisputable reality which precedes the event. For example: "player A won", "it's raining", and "B is talking" are objective antecedents, however, "C is listening" is not, since it can only be objectively observed that "C is silent", but it is not certain that he is listening.

Of course, this definition of context is not perfect, since it does not allow for the influence of culture on the determination of context. Rather, it places context at the same level as the events, outside of culture. In keeping with the model described in section 2.1.1.1, context can be seen as a modifier of the event space coordinates of an event.

This allows for cultural influences to play a role when the individual ascribes meaning to the event. This has the advantage of ascribing each context-event pair a particular interpretation for each cultural basis. It will be shown in Chapter 4 that for the purposes of assessing the meaning of a nonverbal act, such a definition can be adequate. Figure 7 illustrates this definition.

Section 2.4: The importance of display rules

Display rules were mentioned twice earlier in this chapter. Once as modifiers of universal human facial emotion expression (Section 2.3.1.3), and another in Section 2.1.1.2 as the method which is used to assign cultural coordinates to events. While it may no be obvious at first, these interpretations are almost equivalent. This is because, in general, display rules are cultural habits that regulate:

a) that which is appropriate to display, and
b) how to display it in order to achieve the desired communicative effect.

In other words, they are the coding which underlies all nonverbal behavior (Ekman and Friesen, 1969). Of course, display rules apply strongest to deliberate nonverbal behavior, but they can also become so deeply rooted in the sub-conscious of a person that they affect non-deliberate actions as well (Ekman and Friesen, 1969). The very idea that display rules exist, and that they pervade nonverbal discourse is in keeping with the relationship between events and culture which has been proposed.

Display rules contain all of the information necessary to encode a nonverbal act, or any event. This means that once a display rule is known, the appropriate action can be chosen for a given meaning. For example, if an individual wants to express anger, she can "consult" these display rules and find out which nonverbal expression is best suited for this purpose. In abstract terms, she chooses a location in her cultural space which has a non-zero anger coordinate, and executes an event which corresponds to this location. It is precisely her display rules which tell her which event might be appropriate. This is analogous to the task of picking an object in physical space, given a set of directions (2 up, 3 left, and 1 front), using arbitrary units for the sake of example. This is usually done intuitively by most people, according to acquired motor and distance evaluations skills (see Section 2.1.1.2). By analogy, the subject here might choose an event which shows (10 anger, 5 arousal, 10 serious), again, the numbers are in arbitrary units for the sake of example. This is clearly equivalent to the operation described by Equation I (Section 2.1.1.3), but in this case the concept of "display rules" is represented the matrix T_{BE}, where T_{BE} provides the method for encoding the intended communicative event.

Similarly to the previous example, if an individual witnesses an event, he can use his display rules to evaluate it, essentially asking himself: "if I were performing this event, what might it mean"? This corresponds to associating an event with a location in the cultural space: the inverse action. Again, this is similar to the operation described by Equation II (Section 2.1.1.3), where T_{EA} serves to evaluate the event in the receiver's frame of reference. The physical analogy for this action is evaluating the position of an object in the physical frame of reference. These processes occur subconsciously for most

humans (Ekman and Friesen, 1969), but if a computer is to be explicitly programmed to evaluate events, then the processes must be considered from their abstract point of view, making full use of the culture model proposed in this chapter.

The challenge lies in identifying explicit display rules, thus defining a set of correspondences between events and meanings. In mathematical terms, one can write that a function M associates every context-event pair (e, c) with a specific meaning where M(e, c) is a map from the context-event space onto a particular cultural space. This is the same as determining the position P(x, y, z) of an object, where P maps from the object space to the observer's space. According to this vision, the set of display rules for a culture is the set of functions M which enable the evaluation of events in that cultural frame of reference. This issue will be explored further in Chapters 4 and 5.

The existence of display rules implies that the meaning of an act of nonverbal behavior can be deduced once the display rule is known. In other words, this is a vindication that the very idea of translating nonverbal behavior not only legitimate, but also possible. More concretely, the existence of display rules also means that with the correct tool or with the appropriate training, a person from a foreign culture might also be able to decode the original intended meaning. This is what happens to multi-cultural individuals (Section 2.1.2.1). However, if no common rules exist, then a misunderstanding can occur. An example is necessary to illustrate this. Consider Françoise and John, who are French and American respectively. They meet in the street. Upon greeting, Françoise kisses John on the cheeks. This constitutes an event.

According to Françoise's display rules, this event simply means "hello", and is in good agreement with her distant level of acquaintance of John. However, John will start to wonder about Françoise's intentions. According to his display rules, only very closely related or romantically involved people kiss to greet each other. One can see how an unfortunate misunderstanding could follow. The point is that if John knows the French display rules, he can simply evaluate the event according to those rules. If he does not, however, he must either rely on Françoise's knowing American ways, or on some third party to arrange for the misunderstood action to be explained or painted according to American display rules. In any case, an evaluation of the action according to Françoise's rules is necessary before the action can be recast for John's evaluation. This topic will be revisited in Chapter 4.

Section 2.5 Conclusion

This chapter has covered a wide range of topics, starting by giving a definition of culture, and concluding with a discussion of nonverbal behavior and display rules. Both of these discussions were necessary in order to provide a solid basis on which to build the NBCT's frame work.

The NBCT steps which were identified in Chapter 1 (in particular, assigning meaning and translating) were refined according to the new cultural model. This cultural model has the advantage of allowing for a high level of abstraction which is usually

favored by computer scientists. These scientists will eventually design the NBCT, and so they need to be able to understand the scope and importance of translating culturally dependent expressions of information. The new culture model proposed in Section 2.1.1.1 is a concise and mathematically based model which was developed specifically for the purposes of the NBCT. Its mathematical nature should appeal to and answer the needs of engineers far more readily than the many wordy and jargon-based definitions of culture which have been developed by sociologists and ethnographers (Mantovani, 1996).

The discussion on nonverbal behavior attempted to expose the three different types of nonverbal behavior which are important for the NBCT: emblems, illustrators, and affect displays (Section 2.3.1). Moreover, the discussion sought to emphasize the cultural variability of nonverbal behavior, further justifying the need for an NBCT to exist at all. The closing section (Section 2.4) which discussed the topic of display rules, helped to put the entire chapter into perspective, tying the nonverbal behavior theory with the new cultural model. Since the foundation for thinking about translating nonverbal behavior has been laid, it is now possible to discuss some of the specific technological challenges facing an NBCT system. In particular, the Chapter 3 will focus on how nonverbal behavior can be captured and characterize in order to be analyzed.

Chapter 3 Capturing Nonverbal Behavior

I met him at the stone-cutter's, he was taking measurements for posterity.
[Je l'ai rencontré chez un tailleur de pierre, il prenait ses mesures pour la postérité.]

After Jacques Prévert

If nonverbal behavior is to be interpreted and even translated by a computer, it must first be recorded. There are several challenges inherent to the task of recording nonverbal behavior for use with the NBCT. The first question which needs answering is in the realm of psychology: which nonverbal behavior is the most expressive, and which parts of the body perform it. This issue is addressed first, in Section 3.1.

It is then important to recall that the NBCT is meant to be used in business settings, for interpersonal and inter-professional interactions. Therefore, it is an absolute requirement that a non-intrusive method of recording nonverbal behavior be used. By most accounts, the least intrusive tool for recording movements is the video camera because it does not touch the body (Bartlett et al., 1999). However, Picard and Cosier (1997) seemed to opt for wearable computers as the best option. Good results in interpreting facial expressions and body movements using video data have been achieved

by several groups, (Bartlett et al., 1999, Cohn et al. (1999), Yacoob and Davis (1996), Wren and Pentland (1998), Marrin and Picard (1998), Scheirer et al.(1999)). A brief review of these attempts follows, along with a succinct discussion of the consequences of this research for the NBCT. Two sections cover recognition and characterization of nonverbal behavior by computers. Section 3.2 focuses on the face, while Section 3.3 deals with the rest of the body.

Section 3.1 What to track and record

The nonverbal behavior communication channel (Birdwhistell, 1970) can be decomposed into several interrelated sub-channels, each one consisting of a particular body part's movements. Hiltz (Hiltz, 1978) identified some of the important communicative elements which were missing from CSCW tools. In a very crude and yet worthwhile analysis, she ranked "facial expression" as the most important type of "visual information" for bettering communication, placing special emphasis on eye contact, and the rest of body movements as less important. Clearly this analysis is at least partially confused, for while facial expressions are used to produce affect displays (Chapter 2), it is really other parts of the body which are used to perform the emblems and illustrators which were described in Chapter 2. In particular, the head, neck, shoulders, arms, and torso are most often used to perform communicative gestures, even though the feet and leg can sometimes be used (e.g.: tapping one's feet, or shaking one's leg) (Bavelas et al., 1995).

Further, the NBCT is meant to be used in business settings, and more precisely for conferencing. In these situations, when one is seated at a table, the visible part of the body is usually the upper body, and not the legs. Movements of the lower body are not available as part of the face to face communication stream, and therefore there seems to be no valid reason to make a special effort to include them in the capabilities of the NBCT.

It is therefore logical to focus the attention of recording efforts detailed below to the upper body. Special attention should be paid to the face and to affect displays, because of the important of the information provided by this type of nonverbal behavior. However, it is very important to devise an effective method for characterizing emblems and illustrators since these types of nonverbal behavior are the most culturally variable.

Section 3.2 Capturing and Interpreting facial expressions

3.2.1 Technical Considerations

In 1978, Ekman and Friesen devised the Facial Action Coding System (FACS) (Ekman and Friesen, 1978). This system consists of 46 action units, which can be combined in more than 7000 ways to fully describe all possible facial displays and movements. Each action unit or combination of action units corresponds to a particular facial expression, which is not associated with a meaning, but rather are often used to identify emotions in subjects. FACS is completely descriptive (Cohn et al., 1999), and a

different set of tools, called the FACS Interpretative Dictionary (Friesen and Ekman, undated, in Oster et al., 1992), must be used to infer emotional affect from the results of a FACS analysis (Cohn et al., 1999). Besides FACS, there are other methods for analyzing facial displays (as reviewed by Cohn et al.), but only FACS can produce the detail required for accurate analysis of emotions (Cohn et al., 1999). FACS is of particular interest to this study because it allows for the complete, unique, and unambiguous characterization of facial movements.

In an effort to automate the tracking of facial movement, FACS was used by Cohn et al. (Cohn et al., 1999) to train computers to recognize facial displays using feature point tracking. Feature points are points on the face which were chosen because their movement characterizes the overall movement of the eyes, brow, mouth, and nostrils. Cohn et al. used video cameras to record images and track the movements of said feature points on the face of subjects. They found that computer algorithms could identify facial expressions as accurately as trained FACS experts in less than a second by using a commercial desktop 300 MHz computer (Cohn et al., 1999).

There have been other attempts to track facial expressions, using optical flow in particular. Focusing on the motion (or changes) of facial features between expressions rather than on specific points, Yacoob and Davis (1996) were able to devise a data efficient method to track facial movements. Their method is based on "qualitative tracking of principal regions of the face and flow computations at high intensity [of motion] gradient points" (Yacoob and Davis, 1996). This method is detailed enough to

capture blinking, but it requires that movement of the head as a whole be limited. It is also limited to recognizing 6 facial displays, and even the authors realize that more sophisticated capabilities will need to be developed. Bartlett et al. (1999) set out to improve on Yacoob and Davis' method, and they created a new method based on principal component analysis (PCA).

One of the latest methods for facial expression recognition was devised by Bartlett et al. (Bartlett et al., 1999). Using PCA, which is on based the images themselves to track specific facial components rather than on specific points on the face of a subject, they succeeded in using FACS to automatically answer questions such as whether or not one can automatically differentiate truth from lies from the facial expression of a subject. FACS can answer these questions if the analysis is performed by humans, but the challenge was to automate the process. The method involved intensive training and using neural networks. A detailed discussion of how neural networks are trained and used is beyond the scope of this thesis.

3.2.2 Consequences

For this thesis, it is sufficient to understand that, using a video camera, movements of the face can be characterized automatically and accurately into a finite number of action units. The techniques behind this characterization are not perfect at this point, but, as demonstrated above, research is being pursued in this field, and technology is improving. As stated in Section 1.5.2, recognizing and classifying facial expressions is

required if the meaning of these facial expressions is to be extracted (using whichever method seems most appropriate to a designer), and later translated. However, three important issues remain, and these must be addressed before a NBCT can become fully functional. These 3 issues are: a) whether or not images appropriate for analysis can be gotten, b) whether the analysis can be performed sufficiently fast to provide real time feedback to the users, and c) whether or not an initial "base state" can be reliably established for comparison.

All of the methods described in the previous section require that the images be "aligned" –that the subject's face be frontally visible to the camera. This is done artificially when the need arises, and it adds a time overhead to the analysis and interpretation of the pictures. It is also not clear if the search time through neural networks can be reduced below its present state without significant increases in computer power. Because of the synchronization of language and gestures in face to face interactions (Bavelas et al. (1995), Streek (1993)), it is required that the interpretation lag (not to mention the translation lag which will be discussed later) be reduced to a minimum so that the illusion of face to face interaction might be maintained. This will be discussed in greater depth in Chapter 5. Lastly, providing a base state for interpretation using FACS is difficult. This is because few people can be relied upon to pose for a neutral shot which is required for initialization (Birdwhistell (1970), Bartlett et al. (1999)), and in general, trained FACS experts must be used (Cohn et al. (1999), Bartlett et al. (1999)) to produce accurate the neutral faces which are used for training the computer. This problem remains whole in the eyes of the psychology community, and it

is related to the fierce debates concerning self-report of emotions, pain, or affect in psychology experiments in general (Fridlund (1994), Mignault, 1999).

Section 3.3 Capturing and Interpreting Body Movement

As seen in Section 3.2, the problem of recognizing facial expressions is hefty, but it is also being addressed. Whether or not a system such as FACS can be devised for movements of the rest of the body is yet an open debate, but some efforts have been made in that direction. A more basic problem, however, is whether technological solutions are available for capturing and tracking body movements using video input. Clearly, if this is not achieved in both an economical and efficacious manner, there can be no automatic nonverbal behavior cultural translator.

3.3.1 Camera Based Data Collection

The DYNAMAN Model of human motion developed at the MIT Media Lab allows for the full characterization of the movement of the body (Wren and Pentland, 1998) using a digital video camera as the source of input. This system uses powerful interpolation and extrapolation techniques (Kalman Filter and Markhov Chains) to ensure continuous tracking of the subject even in case of partial or full temporary obstruction, or in case of the addition of more subjects to within the camera's field of perception. The drawback to this system is it's expense in computational terms, since it requires the simultaneous use of 4 powerful SGI computers. The system also requires the use of two

or more cameras at the same time (Wren and Pentland, 1998). Lastly, this system does not provide any insight for the decomposition of the human motion into atomic action units, which could be used to recompose any movement, since it performs pixel level analysis rather than taking a holistic approach to characterizing body movement.

Other studies, mostly focusing on the analysis of American Sign Language (ASL), show promise for capturing and characterizing emblems using digital video (emblems are the type of gesture most closely related to sign language (Ekman and Friesen, 1969)). Most noticeably, a project at the MIT Media Lab has succeeded in tracking hand movements using ungloved, unmarked hands. The computer performed with higher than 90% accuracy in analyzing 10 frames per second using a 200 MHz SGI Indy computer (Starner et al., 1996). This is much better than the above methods, but it also focuses on a small part of the body.

As with previous methods, this system uses pixel level analysis and does not attempt to decompose the hands' movements into action units, it rather considers a series of pixel patterns and assigns a corresponding meaning (from a dictionary) to that pattern of pixels. In fact, the system exclusively associates variations in pixel patterns with meanings, no matter how the patterns of pixels are produced. Therefore this algorithm does not directly recognize specific parts of the body, but rather changes in pixel patterns. It is not clear that this system could then be used to identify action units for the body unless specific pixel patterns were associated with action units.

Camera-based tracking and characterizing of movement is still in its formative stages. This will be a potentially serious hurdle for the NBCT, as capturing the movements of the body are absolutely necessary in order to provide an interpretation and a translation of these movements. However, the biggest challenge lies not in the embryonic nature of the technology, by rather in the method used, since it does not allow for easy decomposition of body movements into action units.

3.3.2 Wearable Computers and Sensors

Another promising potential solution is to use wearable computers or physiological sensors embedded in clothing to track and analyze body movements (Marrin and Picard, 1998). The problem with this solution is that it requires a hefty investment in hardware, and that, although the jacket is lightweight, it is more directly intrusive than a video camera, since it is in direct contact with the user. However, the sensors in the jacket can be easily disabled, thus ending the recording of movements. Also, the jacket provides accurate characterization of the intensity of movements, a potentially useful quantity which was not recorded by the video-based systems described above.

The same laboratory has also investigated the use of glasses which can recognize facial expressions (Scheirer et al., 1999). This is much less intrusive than the jacket described above, as it uses a small object which numerous users of the NBCT might already be using. Further, in some cases, wearable computer based methods of data

collection can be more appropriate than cameras as they are more easily disabled without loss of a communicative media. For instance, a user might choose to use a video camera and not to use the translation functions, or want to disable the recording of her movement without sacrificing the visual communication channel. This is not feasible if the camera is used for movement data collection. Wearable computers and sensors imbedded in jewelry, watches, or other common personal objects might be used if these considerations are important for the designers of the NBCT (Scheirer et al., 1999). The downside of these methods of data collection is that they can only gather information about a small area of the body, and so several sensors need to be working simultaneously in order to provide a representation of the total body movement.

At this point it seems that cheap and efficient technological solutions are evolving to solve the problem of tracking human upper body (and eventually total body) motion, even though the tools require ample computer power to perform this task. This means that the NBCT will be able to capture raw movement data which can then be interpreted and translated. However, the problem of characterizing and decomposing these raw data into atomic movements like action units remains unsolved by these methods.

3.3.3 Total Body Action Units

Birdwhistell, (Birdwhistell, 1970), who was interested in decomposing gestures into discrete elements, devised a very crude, if exhaustive, set of notations to characterize all of the possible movements which his subjects might engage in. More precisely, he

focused on the head, face, trunk, shoulders, hands-fingers, hip, legs, ankles, feet, and neck; for each, he identified the set of all possible positions. While these positions do not possess the refinement of the FACS action units, they provide a basis for starting to identify action units for each part of the body.

In its first stages, the NBCT might only focus on specific parts of the body, ignoring those which are commonly not visible during business meetings (e.g.: below a conference table). However, in the long term, developing a FACS-like system for the entire body will be necessary in order to provide the necessary set of discrete action units which will be used in Chapter 4 to assign meanings to nonverbal acts. Such a system would come from psychologists, but it is not clear that the field is currently in need of a total body action unit system.

At this point, and in order to be able to continue with this thesis, it is necessary to assume that such a system will in time be devised, and that it will become usable by designers of body tracking software. In other words, action units describing discrete movements of all elements of the body will be identified. These discrete units, when combined together, will enable the description of all possible human body movement. Of course this is as of yet a hypothetical case, but one which must be realized if the NBCT is to become a reality. When these movements are recorded using any of the methods described above, it will become possible to come up with decompositions of said movements in terms of these future total body action units. It is now appropriate to

consider the assignment of meaning to nonverbal acts by the NBCT, and this is the topic of Chapter 4.

Section 3.4 Conclusion

In this chapter, the technological requirements for recording and characterizing nonverbal behavior were considered. Further, the needed improvements in this technology were briefly considered. Clearly, recording the movement of the human body is necessary in order to be able to describe its meaning and to assign it a translation, and without this step, the rest of the NBCT cannot be developed. In another important consideration, the recording devices must be usable in business environments, and so must be non-intrusive and comfortable. Both camera-based and wearable computer-based solutions were considered, and it is yet unclear which is most appropriate. However, until the use of wearable computers become widespread, it seem unlikely that most business people will feel comfortable using them in business settings, and so the camera-based methods seem to be most appropriate.

As was briefly alluded to in the conclusion to Chapter 1, psychologists and computer scientists must collaborate in devising appropriate methods to simply and effectively capture and characterize nonverbal behavior. A good example of this interdisciplinary collaboration comes from the work of Bartlett et al. (1999), in which computer scientists and psychologists were involved. This collaboration has lead to the use of action units as the basis for capturing and recognizing facial expressions.

However, these efforts have been limited to facial expressions, and a larger problem looms with respect to the characterization of body movement.

The largest remaining hurdle at this last step is the lack of a complete set of action units to describe the movements of the body. Attempts have been made at devising FACS like systems for the whole body, but at this time no reliable such system exists. This is a serious hurdle because action units provide a simple and effective manner for considering nonverbal behavior and for attempting to understand their meaning. The usefulness of action units will be demonstrated in Chapter 4.

Chapter 4 Inferring Meaning

"Custom is almost a second nature."

Plutarch

Once a particular nonverbal behavior has been isolated and identified, the problem of determining its meaning remains yet untouched and unsolved. As was described in Chapter 2, the problem of extracting meaning from a nonverbal act (or any other event) takes place at the level of the individual, using the cultural basis available to that individual. Of course, other individual factors are important, such as the emotional state, mood, or temperament of the person receiving the signal (Picard (1997), Picard (1995)), but while these are significant, they act beyond the scope of the NBCT's present intended capabilities.

It is important to remember that the meaning of an event exists only within a particular cultural reference frame. In general, there will be as many meanings for every performed event as there are people communicating, however similar these meanings might be. In reality, however, there is only one intended meaning (for deliberate events), that of the sender. This meaning, which is the one being studied in this Chapter, is the

meaning which must be recognized in order to proceed further with the NBCT. Recognizing this meaning will allow for the selection of a particular nonverbal act so that it can be represented in the culture of the interlocutor, in other words for the translation of the event.

Section 4.1 How Humans Beings Interpret Nonverbal Behavior to Infer Meaning

The first step in a human's interpretation of an event (after perception) is to discard those events which are "noise", where noise is simply those acts which are not associated with a meaning for a given context and a given culture, e.g.: scratching one's nose or scalp (Birdwhistell, 1970). Of course, events considered noise in some instance might not be in others, and this distinction is very difficult to make (McCarthy, 1996). Scratching the scalp is an example: it can be used to either express puzzlement or nothing at all according to a stereotypical American culture, but it is almost always a very rude act of disrespect according to typical French culture. The ability to discern and dismiss noise events is crucial to ensure that the communicative channel focus on the important, information-carrying, events available to people in conversation. For example, it has been theorized that autistic individuals do not possess the ability to scan and select the outside information which they are receiving, and therefore suffer from information overload. In contrast, healthy humans rely on their emotions to perform this selection (Picard,1997) .

Once communicative nonverbal behavior has been selected and isolated from the noise, the extracting of the meaning can begin. This topic was briefly described in Section 2.4. A brief thought experiment is used to explain how meaning is extracted, making full use of the model for events and culture presented in Chapter 2. Considering two individuals having exactly the same cultural vector basis is akin to considering a single individual performing nonverbal acts in front of a mirror. The meaning which is extracted from the acts by the receiver is precisely that which was intended by the sender. This is because both the encoding of the acts and the decoding of the acts use the same set of display rules (Ekman and Friesen (1969), Montovani, 1996). In other words, to the M decoding function described in Chapter 2 corresponds an M^{-1} inverse function which was used by the sender to encode (or produce) the event within a context. M^{-1} maps from the cultural space to the context-event space, whereas M maps from the context-event space to the cultural space (Figure 8). In Chapter 2, it was explained that the event-context space has the same dimension as the event space. The only difference is that the coordinates of an event in event-space are modified when the event is considered in event-context space. This is in keeping with the idea of the "mega-event" which is described in Sections 4.2.1 and 4.3.1. Finally, the projection of a particular event from one basis to the same basis will be the event itself, without loss of information. In this case, referring to the matrix notation used in Chapter 2 (Equations I and II), M corresponds to T_{EB} and M^{-1} to T_{BE}. It is this last projection which is most relevant at this stage, since translation requires understanding as an absolute prerequisite.

In general, and as was alluded to in Chapter 2, there need not be a one-to-one correspondence between nonverbal acts and meanings (Birdwhistell, 1970), and it is conceivable that M could simultaneously depend on a combination of several (e, c) pairs, i.e.: $M((e_2, c_2) \mid (e_1, c_1))$, where (e_1, c_1) precedes (e_2, c_2) in time. In this sense, nonverbal acts are very much like words and phrases, which can have several meanings depending on the words which surround them. A good example is the difference between a sincere and fake smiles. The context of the smile (as defined in Section 2.3) may not be sufficient to discern between the two, instead the succession of other events which lead to the smile might be more useful.

However, in theory, there are specific display rules which associate particular acts within a context to a particular meaning, very much like a dictionary does for words. From this point of view, the system of rules is deterministic, and M can be known and

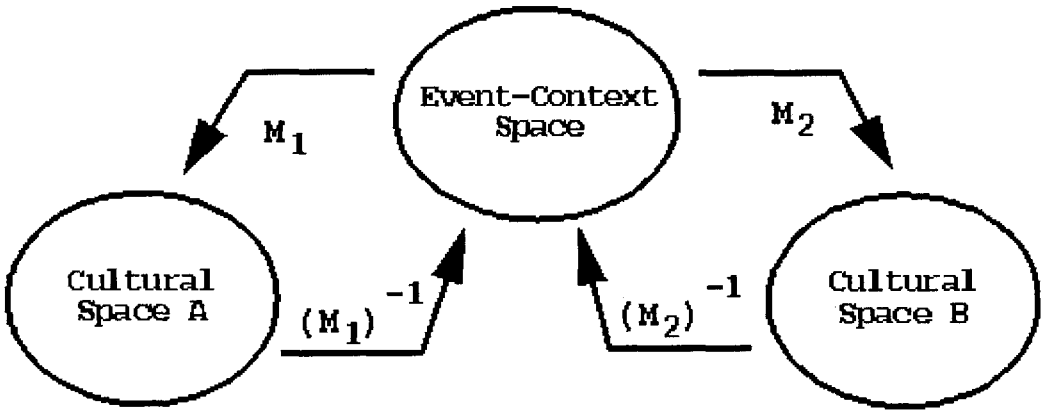

Figure 8. Symbolic representation of the relationships between cultural and event-context space

chosen accurately if sufficient information is available. If the context is unclear, or if the act itself is unclear (e.g.: because the view is partially blocked), then an estimate of the missing information is performed to infer the act or the context, and then the analysis proceeds as above. In this case, the rules of interpretation are still deterministic (M is known), but M might not be chosen accurately. This is because the evaluation of the nature of the act can be probabilistic due to limited evaluation. It is by using these display rules according to this heuristic that humans extract meaning from events, and it is when the rules of two individuals are not entirely similar that misunderstandings can, and do, occur.

According to this culture model, the intended meaning of an event lies within the sender's cultural space. Therefore, in order to understand the intended meaning of an event, the rules of interpretation of the sender must be known. At this stage of the translation process, the cultural space of the receiver is not important because the NBCT is trying to identify the actual meaning of the nonverbal act, from the point of the sender. The cultural space of the receiver becomes important when translating, and this will be considered in Chapter 5. While humans feel that they have intuitive knowledge of the display rules such as M and M^{-1} (Ekman and Friesen (1969), Bavelas et al. (1995)), and while the concepts can transformed into highly abstract rules, it is not clear that display rules can be turned into algorithms which might be easily implemented in computers.

Section 4.2 Applying the Human Model to a Computer System

Assuming that a computer has the capacity to recognize and characterize human movement (as described in Chapter 3, or by using any yet undiscovered method), the problem of ascribing a meaning to this recognized event is the next hurdle in the development of the NBCT. The human model described in Section 4.1, synthesized using the proposed culture model, can provide a very promising framework for considering how to enable computers to interpret nonverbal behavior.

It is first important to remember that not all nonverbal behavior will require translation. However, which nonverbal behavior requires translation cannot be known a priori. Further, all nonverbal behavior will eventually require depiction by the NBCT (Chapter 5). Therefore, it is necessary that all nonverbal acts be associated with a meaning in the frame of reference of the sender so that they might be made available for the receiver.

It is assumed for now that the context of an act can be gotten as described in Chapter 2, from the objective antecedents of the event. A more detailed discussion, presenting an alternative approach from the one offered in Chapter 2, follows in Section 4.3. Given the present assumption, the problem is now to determine whether an M function can be programmed into a computer program. The answer is a priori positive, since in its simplest form, M is only a matching function between context-event pairs and meanings. The rules associating both sides of the equation can be made as arbitrarily simple as possible, leading, of course, to a complete lost of realism.

Research in the field of affective computing, which has striven to associate various human behaviors and physiological signals to emotional state, has encountered numerous pitfalls in coming up with functions like the above M function (Picard,1997). The main problem has been that the rules which associate emotional state to behavior or physiological change are not completely deterministic. Further, these rules depend on numerous, often undetermined, variables (Picard,1997). As was briefly mentioned earlier, this situation is similar to the challenge of associating nonverbal behavior and meaning.

4.2.1 A Deterministic Method

A very simple and crude way to associate a behavior with a meaning is to create an evolving database of event-context pairs, each with a corresponding meaning for a given culture. Each event-context pair in this case would be composed of one of an emblem, illustrator, or affect display, and context. In this case, M is completely deterministic because context-event pairs and meanings are matched in the very design of the database. Also, M applies to a "mega-event" (an event which is taken as a whole rather than decomposed into its parts). The problem is reduced to populating and searching a database of these mega-events. However, it is important to keep in mind that, under these assumptions, given meanings might be expressed with different event-context pairs, while a given event-context pair can have only on corresponding meaning.

This model further requires, and assumes rather naively, that there are a finite number of nonverbal acts which may be performed or have meaning within a culture. There is no guarantee that this fact is either verifiable or true. Moreover, this method removes the flexibility and creativity which are inherent in human nonverbal behavior (Bavelas et al., 1995). For example, in this model, the combination of up and down head shake and hand waving must be considered to be one mega-event, to be placed inside a context, forming an event-context pair. This combination cannot be dynamically separated, or dynamically associated with a smile, a frown, or whichever other act the sender might be engaging in simultaneously. Rather, the act containing the facial expression must be explicitly entered in the database as another, distinct event-context pair. One can see how this would lead to much redundancy as very similar mega-events must be stored individually. If instead, a smile constitutes an individual event, which is dynamically associated with any other event, then the smile needs only to be stored once. A method taking advantage of this idea will be presented in Section 4.2.2.

For emblems, which are highly stylized and very precisely coded (Ekman and Friesen, 1969), the present method might be appropriate. This is because emblems show little variation between repetitions, and because they have explicit definitions. However, there is no way to give the system the flexibility it requires to detect subtle changes in significance of an emblem resulting from its association with a facial expression or an illustrator. Clearly another, more sophisticated, method is needed, not only for emblems, but also for illustrators and affect displays.

4.2.2 A Probabilistic Method

A more efficient, and realistic, strategy is to separate the incoming nonverbal behavior signal into action units. Action units were described in Chapter 3, and can be used to decompose nonverbal behavior into atomic elements. Action units can then be recombined dynamically into acts. Meanwhile, a probabilistic method can be used to

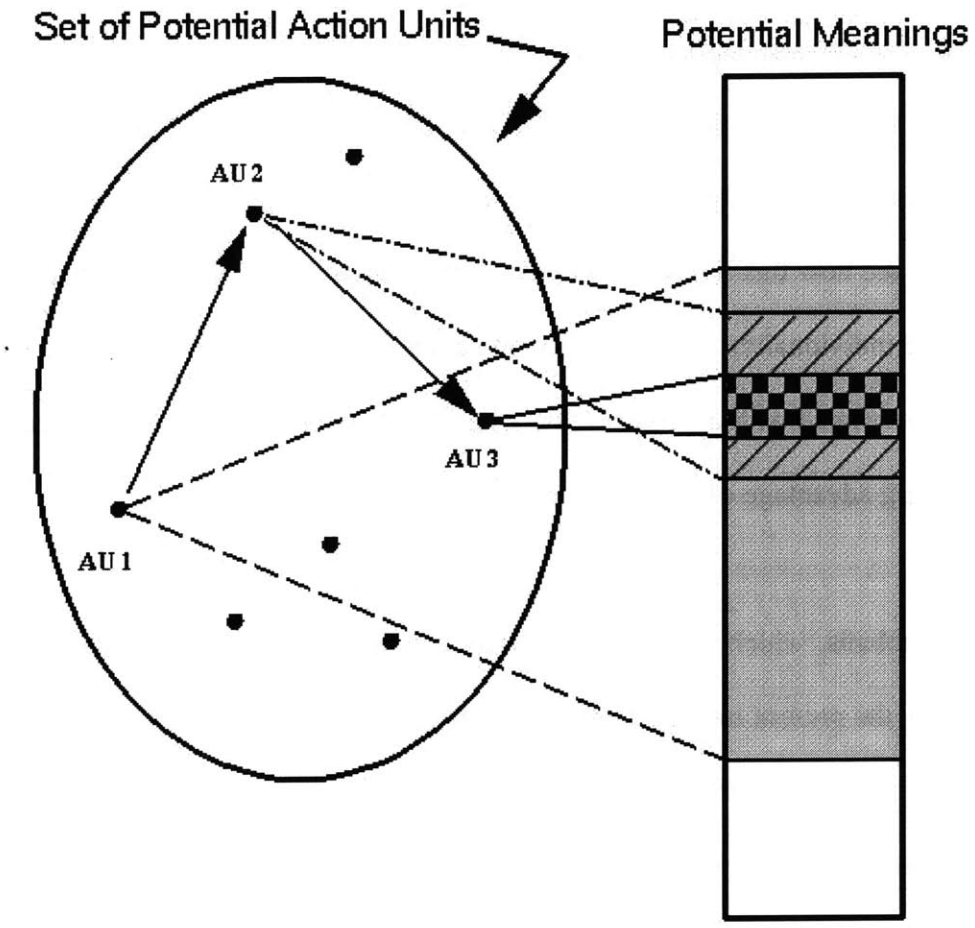

Figure 9. How meaning is assigned by the probabilistic method. The arrows indicate successive action units, the number of potential meanings is reduced after each successive action unit

reduce the possible number of meanings which is associated with each action unit, and with each combination of action units. In this case, M becomes a function which looks

for the most likely meaning, given a succession of action units, as shown in Figure 9. If no single meaning can be found for the particular sequence of action units, then a new meaning must be created from the elements which could not be reconciled. Loosely speaking, this allows meanings like "sad but lying" to be formed from "sad" and "lying". The culture model, which allowed for the decomposition of meanings into independent cultural vectors is hereby put to use. Without such a model, it would not be possible to decompose the meaning of events, and this method would be worthless.

This is a very significant improvement over the method from Section 4.2.1 because it allows for the possibility of contradictory meanings of individual actions weighing in on the final meaning of the overall nonverbal behavior. Also, the system is not restricted in the number of meanings which it can recognize. For instance, this model can differentiate between the "come here" hand emblem (fist with index finger curling and uncurling) associated with a friendly smile and the "come here" emblem associated with an evil smile, and assign a different meaning to each.

More importantly, this method gets rid of the notion of the act (emblem, illustrator, or affect display) as the element of nonverbal behavior, and rather focuses on the succession of action units which make up an act. In this sense it is like focusing on the potential meaning of each word as it is being uttered in attempting to ascribe a meaning to an entire sentence. In this manner, each combination of action units, rather than a whole act, constitutes an event, this is in keeping with Birdwhistell (1970)'s ideas about communication.

The real advantage of this method is that it is flexible, in that it allows unforeseen or extremely complicated combinations of action units to be recognized. Describing the mathematics which would be required to implement such a model is beyond the scope of this thesis, but one can see that the algorithm requires training in order to function. The method works on the probabilistic matching of action units (or combinations of action units) with meanings, and these probabilities need to be fed into the system in some way. This is the same problem as filling a database, but if a learning algorithm such as neural networks is employed, then only data are actually required, and with extensive training, the network can become more dynamically adaptive than a traditional database (Blumberg, 1997).

This method is expected to be able to deal with unexpected nonverbal behavior, and in this sense, it is clearly superior to the first method previously outlined. The creativity of human beings in producing nonverbal behavior –especially illustrators (Ekman and Friesen (1969), Birdwhistell (1970))– requires that methods of interpretation be flexible in order to provide useful information to a translator or a displaying system. This makes the second method preferable over the first one. However, nothing has been said about context, as it was assumed from the start that context was simply determined from antecedent actions. A close look at the difficulties of determining the context of nonverbal behavior is warranted at this point.

Section 4.3 The Challenge of Context

This section recalls the previous definition of context (Section 4.3.1), and presents a new manner of approaching the problem (Section 4.3.2). This new method is better suited than the old one for use with the interpretation model presented in Section 4.2.2.

4.3.1 Previous Definition of Context

Determining the context of a nonverbal action is difficult. Several non-related elements make up context, and only a few can be available to the NBCT at any given time. This was alluded to in Section 2.3.2. There, it was argued that context should be determined from objective antecedents, and that context-event pairs should be used for evaluating meaning. In this sense, the context and event became blurred (Figure 2). It was the same in the first interpretation method proposed in Section 4.2.1, where each context-event pair formed a whole, termed "mega-event", and was ascribed a specific meaning.

4.3.2 A New Approach

The second method for extracting meaning (Section 4.2.2), however, paid little attention to the context of a particular action unit. Instead, it focused on the range of potential meanings of that action unit. Clearly, putting an action unit within a context can help reduce the range of possible meanings, and therefore make the method more effective. In this approach, the historical elements which are of particular interest to the

NBCT include the objective antecedents, but also the previous nonverbal acts (which in general cannot be considered to be objective antecedents). Both of these are events in their own right, and context can be seen as a function of an ordered series of relevant events. If one considers action units to be events, then context is a function of a series of action units. Action units are the lowest level of granularity in human behavior, indeed this is why Ekman and Friesen (1969) developed them in the first place. Therefore, this method guarantees that no further refinement will be possible (at least with respect to the level of observation of human movement).

The field of natural language processing is very much concerned with finding the context of words, so that it can differentiate between homonyms or alike words and phrases (e.g.: a bout [noun] and about [adverb]) (Williams, 1999). This is not unlike the problem at hand, although in the NBCT's case the question is not to differentiate between the individual elements of the sentences (the words or action units) but rather between the potential meanings of these elements or of the whole sentence. Assuming that the probabilistic method presented above is employed for ascribing meaning, the context of each action unit must be determined. A single action unit (AU) is preceded and followed by several others, in the following manner:

...,PrevAU2 , PrevAU1, **AU**, NextAU1, NextAU2,...

This sequence can make up a complete or partial nonverbal act. In order to determine the meaning of the overall sequence, we need to find the potential meanings of each

individual action unit (as described in Section 4.2.2). The notion of determining context can be reduced to evaluating the probability that AU has a meaning M(AU) given a previous and next AU:

$$P (M(\mathbf{AU}) \mid PrevAU1, PrevAU2) \text{ (Williams, 1999)}$$

If necessary, M(AU) can depend on more AU's, even those following the AU in question, and the meanings of these AU's can be determined in a similar manner. This is simpler (and so potentially more desirable) than the context models previously proposed, but it assumes that nonverbal behavior behaves like spoken language, with a syntax that can be explicitly determined. However, that fact is heavily debated (Bavelas et al. (1995), Birdwhistell (1970)) and the point of this short introduction to context is not to resolve this issue. Rather, the point is to introduce a manner of thinking about context so that this problem might be further explored.

Section 4.4 Conclusion

Having used the preceding heuristic to deal with context in extracting the meaning of a nonverbal acts, the NBCT possesses a definition of said nonverbal act. This is the last required step before a translation can be proposed. In any two-language dictionary, no definitions are provided. Instead the user is expected to use her own language's dictionary to find definitions (or meanings), the two-language dictionary simply presents a version of that meaning which is understandable to the user, both in term of the

language used, and of the symbols used to express it: the meaning is the same. In the coming final stage, methods will be explored to allow the NBCT find appropriate translations of nonverbal acts, and equally appropriate symbolic representations of these translations. This is covered in Chapter 5.

More importantly for the overall task of this thesis, this chapter has discussed the need to evaluate the meaning of communicative acts at the level of the sender. This is crucial to ensure effective communications in the workplace. If co-workers are to trust each other, as is the goal of the NBCT, then the intention of the sender should be taken into consideration. This is intended to prevent the types of misunderstandings which are the plague of email systems (e.g.: unwarranted email wars and flaming (Picard and Cosier, 1997)), and to make sure that intended criticism is received unambiguously.

Chapter 5 Translating and Representing Nonverbal Behavior

If Cleopatra's nose had been a little shorter, the whole face of the world would have been changed.

Blaise Pascal

Assuming that the meaning of a particular nonverbal behavior can be established in the cultural frame of reference of the originator (as was described in Chapter 4), an appropriate corresponding nonverbal act must be chosen for display to the recipient. The recipient should then be presented with this new nonverbal act in the most intuitive and evocative manner. Once these last two steps are completed, the translation process will then be completed, and the NBCT will be ready.

This chapter is composed of two main sections. In the first section (5.1), there is a brief description a very simple heuristic method which might be used to associate meanings with nonverbal acts. In other words, the task is to perform the reverse of the actions discussed in Chapter 4. Later, in Section 5.2, this chapter explores the challenges and requirements of displaying translated nonverbal acts so to the recipient. Unambiguous displaying of translated nonverbal acts is important. This is because

NBCT users will be exposed to these displays rather than to their interlocutors during NBCT-supported interactions.

Section 5.1: Translating a Nonverbal Act

5.1.1 Context

Fortunately, the problem of choosing an appropriate behavior to correspond to a meaning is simpler than its inverse. There are several reasons for this, and none is more important than the absence of context considerations. This is because context as it has been described in this thesis (both in Chapter 2 and in Chapter 4) is relevant for ascribing meaning, rather than for choosing a particular nonverbal act. Of course, in face to face communications, context does determine which words or nonverbal acts are appropriate (and performed), and which are inappropriate (and avoided). However, this is only true in the vaguest sense of the word context. In the case of the NBCT, differences between "business context" and "casual context", for example, need not exist. This is because, provided that the context in which the NBCT will be used is pre-determined, all inappropriate behavior can be excluded from the potential choices. In fact, this context is predetermined, as outline in the requirements: the NBCT is meant to be used for interpersonal and inter professional communication. This is the easiest way to address this issue, and it results in a direct and potentially simple method for matching meanings with nonverbal acts.

5.1.2 Translation

The transformation of a meaning into a representative nonverbal act is simpler than applying display rules (M) to a nonverbal act to find its meaning (Chapter 4). This is because it is easy to limit the output of the NBCT to a finite number of expressions, even if that number remains large in order to ensure realism. In the previous chapter, the challenge of inferring meaning lay in the multiple possible meanings which can correspond to an event, especially when this event is reduced to the level of action units. However, while it is possible to express the same meaning in a variety of manners, the NBCT is not required to accommodate such flexibility. This is because a single meaning may not arise frequently enough to cause the unnatural repetition of particular patterns by the sender's representation. If this were to become a problem, further refinements could be achieved by providing a list of possible equivalent, different nonverbal behaviors from which a random act could be chosen: this would provide variety. Of course, this paradigm assumes that there are a finite number of meanings which can be expressed. Further, it assumes that since meanings exist outside of cultural bases, all meanings are at least theoretically visible from all bases.

The first assumption is not verifiable and most likely erroneous, moreover, it does not allow for much flexibility in exchanging nonverbal communication. This was already a specific problem with the previous chapter's deterministic method of assigning meaning to acts. However, in this case, it can be argued that the number of meanings must be less than the number of possible combinations of action units, since several combinations can have the same meaning. Moreover, even the probabilistic method of Chapter 4 made use

of an instantaneously finite (if ever increasing) number of meanings to choose from, and so at this point it is not clear that it is beneficial to consider the possibilities and complications of the existence of an infinite number of meanings for the NBCT.

The last assumption above is in keeping with the culture model outlined in Chapter 2. However, the culture model contained the important caveat that an event might have an empty projection onto a particular culture (in other words, it has no attached meaning). This assumption of the translation model requires a clever solution and it highlights significant problem in cross cultural communication which must be addressed by the NBCT.

If an event has no attached meaning for a particular individual, then that individual may not even be aware of the fact that information has been sent to her (irrespective of the communication channel used). For example, if an Inuit tried to describe snow to a Pygmy, it is unlikely that the Pygmy would even understand the concept of frozen water, let alone of frozen precipitation (provided, of course, that the Pygmy has not been exposed to ice before). In other words, this can occur if a meaning is simply never considered in a particular culture, and therefore has no representation in that culture, because there are no basis vectors provide a decomposition of the meaning. It is important to understand that this is different from a simple misunderstanding; rather, it constitutes a complete loss of information. This can cause dire problems, as information which was thought to have been provided is never received (Gercik, 1996), and corresponding answers or acknowledgements are never sent.

The NBCT can help alleviate this problem, making sure that no information is lost between two communicating parties. Since a meaning is always associated with a nonverbal act by the NBCT (as described in Chapter 4), it does not loose information. The NBCT can then make sure that said information is conveyed to the intended recipient. For this very reason, it is necessary that the NBCT be made capable of expressing all meanings in all cultural spaces which might be using it. Clearly, the above assumption points to a necessary characteristic of the NBCT. In fine, the NBCT will be required to make approximations, or possibly to explain in words the sender's meaning to the recipient. The issue of representation of translated nonverbal behavior is the topic of the rest of this chapter.

Section 5.2 Representation of Translated Nonverbal Behavior

Once a nonverbal behavior has been associated with a particular meaning through the translation process, the last remaining hurdle is to find an efficient and communicative way to provide a representation of that nonverbal behavior to the intended receiver. There are two important issues here. The first is that of coordinating nonverbal behavior with language (Section 5.2.1), and the second of picking the most appropriate representation technique for the purposes of the NBCT (Section 5.2.2).

5.2.1 Coordination of Nonverbal Behavior and Spoken Language

Before the issue of how nonverbal behavior feedback should be provided to the users of the CSCW tools through the NBCT, it is important to explore the timing of that feedback with respect to speech. In the case of the NBCT, speech can be either written text (in chat rooms, for example) or actual voice, depending on the functionality supported by the CSCW tool in use. When face to face communication occurs, gesture and verbal communication are combined into a single, integrated information stream (Vilhjálmsson and Cassel, 1998). As was already discussed previously, the division between nonverbal and verbal communications is artificial (Birdwhistell (1970), Goldin-Weaver, Bavelas et al (1995), Streek (1993)). Studies of the coordination of nonverbal acts with spoken language have attempted to establish rules concerning the synchronization of speech and gesture. However, success has been limited, caused in part to significant divergences in the methods of study (Streek, 1993). While debates among psychologists are not relevant to the NBCT, the discovery of such rules would make the implementation of the display mechanisms much easier.

The common, if intuitively obvious, result of all studies, is that the coordination of gesture and language does matter, in particular for affect displays and illustrators (Bavelas et al., 1995). However, it is not possible to suggest a sophisticated algorithm because there does not exist a set of well documented rules to base the algorithm on. Instead, the following simple method might be used, making use of the fact that both the audio channel and video channel record information while the NBCT is in use. Individual action units, once they are decoded, can be matched with individual time

stamps in the speech flow. Provided that the speech flow output to the receiver is delayed until the nonverbal acts is translated, the two channels can be matched again at the output. This is neither an elegant nor an efficient solution, but is has the merit of addressing this very important matter in a simple way. Furthermore, a very similar method was used by Vilhjálmsson and Cassel (1998), in their BodyChat interface, to match chat text and avatar movement (Vilhjálmsson, 1999).

5.2.2 Displaying Nonverbal Behavior

Nonverbal behavior is by nature a very rich and subtle media (Birdwhistell (1970), Bavelas et al (1995)). Deciding what to represent and how to represent it are crucial issues which must be addressed. There are several options available to provide representations of the translated nonverbal acts. The NBCT will limit itself to two dimensional representations of users because issues of maneuvering, and interacting in three dimensional space constitute a serious research topic onto themselves (Vilhjálmsson and Cassel, 1998). The two dimensional options include: symbolic descriptions, text descriptions, still frame animation, continuous animation, and video emulation. For each method, the user will be represented by an avatar, or "incarnation or embodiment of a person" (Webster, 1981). Avatars may be life-like, abstract, or fantastic. They have become a common fixture in chat rooms on the World Wide Web (Microsoft Corp (1999), Vilhjálmsson and Cassel (1998)). The merits of each of these methods with respect to the issues identified above will be discussed below, recalling the specific intended forum for the NBCT, namely professional settings.

5.2.2.1 Symbolic Descriptions

In this context, symbolic descriptions are defined as representations of particular ideas, events, or places, which make use of abstract, encoded symbols rather than of concrete pictures. For example, written alphabet-based language can provide symbolic descriptions, as can emoticons or pictograms. It is possible to describe nonverbal translated behavior using words or symbols (Streek, 1993), and such a solution should be considered carefully because it has a very low overhead in terms of displaying costs.

While it is true that a pictorial depiction of any concept can be more revealing than word or symbolic descriptions, these have the advantage of being very efficient and highly synthetic. These are both important characteristics of displaying techniques, as will be discussed later, in the sections concerning animation (Section 5.2.2.3 to Section 5.2.2.5). In short, very simple symbols can be used to convey a specified set of meanings unambiguously. They also can be combined to express more complicated meanings, much like emoticons or pictograms, remaining unambiguous because of their simplicity. Ambiguous descriptions of nonverbal behavior pose a serious threat to the entire NBCT endeavor because they will confuse users further than the absence of descriptions. Another advantage is that symbols require low level of computer power for displaying, since they are simple and abstract, as compared with some of the later options (they make use of the ASCII text format, and thus requiring less transmission time or storage space).

Unfortunately, symbolic representations require users to learn the semantics of the symbols in use, and symbols are by nature limiting, as is written language itself. Using this type of representation is like qualifying language with gestures, which has been dismissed as an inaccurate model (Birdwhistell, 1970). It does not provide a real life-feel, and it most likely cannot capture the many subtle variations of meanings which are inherent in nonverbal behavior because it approaches written language in its nature (Picard and Cosier, 1997). Symbolic description, like emoticons, has its place in nonverbal communications (Fridlund, 1994), but it is unlikely that it provides an acceptable solution by itself; rather it must be associated with other types of representations. Symbolic representations can be useful if a translation described in another manner is unclear for the reasons outlined in Chapter 4 (the meaning does not exist in the recipient's cultural frame of reference). In this case in particular, symbolic representations can be useful to clarify the translation.

5.2.2.2 Text Descriptions

Text descriptions are more sophisticated than symbolic descriptions, and they require less learning on the part of the user. This is because written language is a common tool for expressing ideas which most potential NBCT users would already be familiar with. This method also allows for detailed explanations of meanings, in the event that no nonverbal behavior exists in the interlocutor's culture to convey the intended meaning (see Section 5.1.2 above for an example).

However, allowing users to escape the norms and limitations of purely written communication is one of the goals of CSCW tools (Mark (1998), Boyer et al. (1998), Adams and Toomey (1998), Hiltz (1978)), and, by extension, of the NBCT. Moreover, one must wonder about the ability of users to simultaneously listen to an interlocutor and read a necessarily verbose description of the nonverbal behavior. Judging from the literature on gestures, it is very difficult to give detailed and communicative written descriptions of nonverbal behavior. While text description does provide possibilities for detailed feedback and precise portrayal of nonverbal behavior, it is a self defeating representation method for the NBCT, and therefore does not provide a viable option.

5.2.2.3 Still Frame Animation

In this paper, still frame animation is taken to denote the successive display of discrete images showing different poses. The goal is not to provide cartoon-like animation, but rather a "comic strip" of the nonverbal behavior of the sender. This requires that an embodied representation (abstract avatars, drawings, or photographs), and a database of possible nonverbal states (action stills) of the users be available. After a translation is completed, a single representation can be chosen from the database. There are several advantages to this method of displaying nonverbal behavior over the two previously mentioned. In particular, provided that the embodied representation is life-like, then this method approximates face to face meetings better than the previous methods because it provides the illusion that the actual interlocutor is present (Picard, 1995).

Experts comic strip artists can be consulted to devise the most efficient ways to utilize this representation. There are inherent challenges to the discrete or animated

Figure 10. Embarrassment and disgust. (Walterson, 1988)

representation of nonverbal behavior. For discrete animation, each frame must display a clear and unambiguous set of action units, corresponding to a meaning, as exemplified in figure 10. Most studies in this field have focused on how the emotion of the represented character are conveyed through drawings (Uderzo (1985), Thomas and Johnston (1981)). The challenge of displaying nonverbal acts for the NBCT is slightly different, since the emphasis is not on what the avatar is feeling, but rather on what it is doing (i.e.: the user is acting embarrassed, even though he might not be). However since most people have become unconsciously familiar with the methods and shortcuts used by comic strip artists to coordinate language and action frames (through repeated exposure to this form of media) (Picard, 1999), it seems logical that such methods could be leveraged for this endeavor. Clearly the principle disadvantage of this method in terms of realism is the static nature of the display. A more animated representation would provide a more life-like experience.

5.2.2.4 Continuous Animation

Continuous animation is simply cartoon-like animation. In this type of animation, a non-photographic avatar of the users is animated in real time. This is the most promising of the methods outlined so far because it provides the most real-life feel (Vilhjálmsson and Cassel, 1998), and because it consumes very low bandwidth as well (Stroud, 1999). It is also the most challenging (along with video emulation) of the methods for two principal reasons. First, the animation itself requires a complex set of techniques, software, and expertise, and second, the timing of the nonverbal acts with respect to speech becomes crucial. This latter difficulty is the most important, and will be discussed first.

With real-time representation, the issue of the exact correspondence of a nonverbal act with its associated verbal input becomes central. This is because humans are used to seeing coordinated verbal and nonverbal acts both in real life (Streek, 1993), and in animations (Thomas and Johnston, 1981). NBCT users will therefore expect that a sophisticated system using animation will provide them with such coordination. However, as was explained earlier, there are no explicit rules to determine how and when motion and words might be associated.

In static representation systems, the coordination might be approximate, with the user (or reader in the case of comics) mentally "plugging-in" the still action at the appropriate moment in the speech associated with each frame (Uderzo, 1985). In real-time, animated video, the viewer expects the association to be made for him or her

(Thomas and Johnston, 1981). This means that either the simple heuristic described above (noting the coordination of language and action units in the input stream and maintaining that coordination in the translated output) must be used for this type of representation of nonverbal acts, or a more sophisticated algorithm must be developed. The existence of such a sophisticated algorithm remains purely speculative at this point, and so the focus should remain on perfecting simple heuristics instead. The current heuristic's most profound weakness is that it assumes that the coordination of speech and gesture is the same across cultures. There is strong evidence that this is not the case (Streek, 1993), especially for illustrators, which serve to emphasize, illustrate, and convey different parts of a particular sentence. However, Streek (1993) found that in general, illustrators used to emphasize particular words tend to occur immediately before the word itself is spoken, serving as kind of flag to warn the interlocutor of the importance of the impending speech. This rule was hard wired into the BodyChat system of Vilhjálmsson and Cassel (1998), and words were properly emphasized with head nods and movements. However, this rule is not universal, and should not be considered to be, but it shows the potential for discovering possible rules which might aid in the elaboration of more potent algorithms.

The first difficulty mentioned above –namely the technical challenge of producing quality and life-like credible animated representations of NBCT users– has been studied extensively by animation artists in the realization of their craft (Thomas and Johnston, 1981). For the specific needs of the NBCT, animations will need to be created in real time, from the input provided by the motion recognition software and the translator.

According to Robert Jensen, animator at Pixar, Inc., this technology exists (real time rendering), has been demonstrated (Jensen (1999), Picard and Cosier (1997)), and is been considered an appropriate solution for videoconferencing over low-bandwidth networks (Stroud, 1999). There are also drawing limitations imposed by the very nature of animation. Thomas and Johnston (1981), claim that no more than one emotion (or meaning, in the case of the NBCT) can be displayed at a single time by a single character. In other words, displaying techniques must synthesize and simplify all behavior in order to ensure proper understanding from the intended audience. Thomas and Johnston (1981) also claim that methods which use highly synthetic notations present the least potential for misunderstandings, provided that the codes are known to users. It is necessary to consider this fact choosing and animating an avatar to avoid confusing the audience and to ensure that the message intended by the animator (and by extension the sending user) is clear.

The NBCT is not limited to accurately representing the emotional state of the user, but his or her physical movements in general. These movements are made up of several concurrent action units, and their representation requires portraying several action units simultaneously. Moreover, these action units must be carefully selected and exaggerated (Thomas and Johnston, 1981) to secure understanding of the meaning. In other words, the NBCT will be required to make use of the most advanced animating techniques available to foster a sense of realism in the communicative interface.

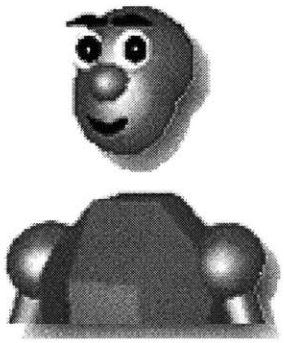

Figure 11. An abstract avatar (Vilhjálmsson and Cassel, 1998)

The advantage of this method over video transmission (independent of translation) is that the level of precision necessary to convey the requisite information is much lower. This is because the animator can decide how abstract the avatar of a user can be. As explained above, the avatar can be extremely simple without significant loss of information, and thus the image can require very low power to generate. A good example of such an abstract avatar is presented in Figure 10 and 11. Video emulation is the last representation method for translated nonverbal behavior which will be discussed in this thesis. The next also section addresses the potential problems arising from using cartoon-like avatars in business settings.

5.2.2.5 Video Emulation

The concept of video emulation makes use of a technique called texture mapping. Using this method, photographs of users are animated using coordinate changes to stretch images. This, if done correctly, can provide the most realistic representation available for

use with the NBCT. Live video is not usable in this context, since by definition is it not translated (i.e.: it is live). The concept of video emulation is simple. Starting with an image, a three dimensional representation of a user can be created (Palisades Research (1999), Bourke (1999)), and this 3D avatar can be animated either in a 2D or 3D environment, as required.

In this sense, video emulation is the same as continuous animation, but it is done with photographs rather than graphical avatars of users. Thus, video emulation faces the same challenges of gesture-speech coordination as continuous animation. However, there are no questions as to how the user should be represented, since a photograph is used. This is both a simplifying factor, since it removes a degree of freedom and uncertainty from the NBCT, and a complicating factor, since it does not allow for abstract representations of users to serve as their representation. This latter point is significant since by reducing the avatar of the users to simple but effective drawings, one can reduce the power and time needed to render an animation. The use of real photographs however, is a significant improvement, since it limits the ambiguity which might arise from using non photographic representations of users (Picard and Cosier, 1997).

While numerous virtual environments promote the use of graphical avatars to represent users (e.g.: Microsoft Comic Chat (Microsoft Corp, 1999), BodyChat (Vilhjálmsson and Cassel, 1998), and WorldChat (Worlds Inc., 1999), to name only a few), these non-photographic representations are usually used in ludic (i.e.: non-professional) environments. Since the NBCT is intended to be used in business settings,

graphical representations may not be sufficiently "serious" to be used by the NBCT. Adams and Toomey (1998) found that "businesslike" avatars such as photographs were most often chosen for formal interactions, while "more individualized avatars [were used] during informal interactions". It is unquestionable that an animated photograph of a business-suited user looks more professional than a cartoon-like rendition of that person. The NBCT could then provide at least two sets of representations for each user. The first would be used for formal interactions, and require using video emulation. The others, possibly customized by the user himself, would be used in informal interactions, and might make use of any animation method (still, dynamic, or video emulation), depending on the nature of the representation and of the interaction. Indeed, the very choice of the second avatar could be a way for NBCT users to express something about themselves. For example, in Adams and Toomey's (1998) experiment, some users chose to be embodied by pictures of domestic animals, thus sparking discussions about their pets.

Section 5.3 Conclusion

This chapter has focused on two problems, two of the important remaining issues as concerns the NCBT, namely translating nonverbal behavior, and representing that behavior. The first issue was reduced to simple matching problem by the preliminary work done in Chapters 2 and 4. However, the second issue was spiny and currently does not have a final complete solution.

Keeping in mind the intended business function of the NBCT, it is important to realize that neither the translation nor the representation scheme can afford to make mistakes. The representation scheme must also provide credible and respectable avatars for the users. Clearly, what constitutes a respectable avatar might be culturally specific, and the NBCT could assist users in choosing such an avatar, depending on the intended interlocutor. Using video emulation (Section 5.2.2.5) seems to be the surest way to create the required sense of realism and respectability for the avatars and their users.

Animated representations in general also seem better than static representations because they provide a more life-like experience for the user. Remembering Walther's (1995) finding that most business people reserve the discussion of sensitive issues for face to face meetings, the NBCT should aim to provide a communication experience as close as possible to a face to face encounter. The methods outlined in this chapter aim to provide a "first-order" solution to this very complex problem. A short review of the contents of this thesis is provided in the Conclusion (Chapter 6), followed by a short discussion of some further considerations on the topic of the NBCT.

Chapter 6 Conclusion

"The most important thing about having goals is having one"

Geoffrey F. Abert

The conclusion of this thesis is made up of two parts. In Section 6.1, a review of the contents of the thesis is given. In Section 6.2, some important and yet not discussed topics are briefly introduced.

6.1 Brief review of the contents of the thesis

The overall goal of this thesis was to outline the motivation and requirements for a nonverbal behavior cultural translator (NBCT) to be used in business environments. This thesis was also aimed to serve as a stepping stone towards future research in the field of computer supported cross-cultural collaboration. This thesis focused specifically on the need for nonverbal behavior information exchange in distributed communication environments, and also on the specific difficulties which potential designers of an NBCT might face. In order to achieve the goals of the thesis, several steps were taken. These steps are described in the following paragraphs.

First, the motivations and need for cross-cultural support in computer supported collaborative work (CSCW) tools were explored in Chapter 1. Chapter 1 also detailed the importance of nonverbal behavior in communications. In fact, it is argued that nonverbal behavior is a necessary part of effective communications where effective communications are those communications during which minimal information is lost. The development of effective communications was identified as a crucial prerequisite in building "association based" trust.

Developing identification-based trust between co-workers is important because it ensures that they will have the same goals, at least with respect to the company. In Chapter 1 the argument was made that the issue of nonverbal communication is tied with the issue of effective communication. This last issue is in turn tied to identification-based trust, which is one of the major cruxes of the entire CSCW paradigm. Indeed, establishing trust requires effective communication and interpersonal interaction, and CSCW tools are notoriously poor at supporting either. Chapter 1 concluded that in order to develop this identification-based trust, the NBCT must support nonverbal communications, and must translate this nonverbal behavior as necessary between cultures.

Chapter 2 sought to establish the cultural variability of nonverbal behavior, along with the cultural dependence of meaning. For this purpose, a powerful and concise definition of culture was provided. Culture was defined as a vector basis, and it included

display rules as a means of encoding and decoding nonverbal behavior within that basis. This definition of culture enabled a convenient and abstract formulation of the task facing the NBCT.

In particular, the details of the interpretation of the meaning of an event were investigated. It was found that an event can have several meanings: in general one for each of those witnessing it. Thus, before the event can be translated, its meaning must be assessed according to the cultural basis of the originator. This ensures that the original meaning is known. Clearly, these steps are complicated and require that nonverbal behavior be captured and characterized. A brief introduction to this topic was offered in Chapter 3.

Technological solutions to the problem of recording and classifying human movement are currently being developed by teams of computer scientists and psychologists. These solutions are still in their infancy stages, and require much development. However, automating the FACS method (Ekman and Friesen, 1978), some teams have been able to discretise human facial movement into elemental components: so-called action units. These action units can be combined to make up all facial movements and expressions. Unfortunately these action units only cover the face, and do not extend to the rest of the body. The methods of nonverbal behavior which were presented in Chapter 4 make the implicit assumption that action units will be developed for the entire body. This is an important caveat, but it was a necessary assumption in order to proceed with the further steps of the NBCT.

Chapter 4 proposed two separate methods for inferring meaning from nonverbal behavior. The first one is deterministic and simplistic. It is also very inefficient because it requires that every possible event be explicitly described and associated with a meaning. This contradicts the inherent flexibility which humans demonstrate when using nonverbal behavior, often freely associating gestures to convey new or modified meanings. However, this method made use of a simple model for context, essentially considering that an event and its context form a mega-event, which may not be separated.

The second method attempted to mimic the flexibility of expression which humans can demonstrate. This method is based on the probabilistic association of meaning with a particular sequence of acts, and it makes use of the decomposition of nonverbal behavior into elemental action units, and considers the possible meanings of these action units. This newer method requires a more sophisticated definition of context, which was provided. Briefly, this new definition of context suggests using the fact that a particular sequence of action units can affect the meaning of each individual action unit. Thus, the meaning of a particular action unit is probabilistically related to the nature and presence of its neighbors, much like the meaning of a word is affected by its surrounding words.

It is important to realize that the meaning of a nonverbal act must be evaluated at the level of the originator of that act. This can be used as a guiding rule in cross cultural communication, whether computer supported or not. This would help avoid

misunderstandings as individuals ponder what their interlocutor really meant, rather than what they thought they understood. Listening (and understanding) is a basic step in establishing effective communication.

Once a nonverbal act has been associated with a meaning in the frame of reference of the sender, the display rules of the receiver can be used to create a new nonverbal act with the same meaning. Chapter 5 briefly describes this process. Given a new, translated, nonverbal act, the NBCT must provide a useful and intuitively communicative representation of that act to the users. The representation of the act must also be acceptable in a business environment, both in the nature of the avatar and in the readiness of the information available. Moreover, the nonverbal act must be coordinated with the other communication channels supported by the CSCW tool (e.g.: speech, or text) to provide a life-like communication experience. This poses a stiff challenge as the capture, interpretation, and translation of nonverbal acts inevitably takes more time than the mere transfer of either sound or text. However, animators have studied in depth the requirements of realistic animations, and some solutions are being devised by the creators of chat environments. At this time, video emulation based on photographs seems the most appropriate method of nonverbal behavior information transmission, especially in business settings. This is because traditional business settings and meetings are often formally organized, with strict rules and dress codes, and so formal-looking, life-like avatars are required. It is conceivable that for more informal interactions, co-workers could select fantastic avatars to represent them. In general, the NBCT should give users

the freedom to choose their avatars (or even advise them in their choice) in order to suit the multiple possible types of interactions which can occur in a distributed workplace.

Each of the chapters of this thesis highlighted a different aspect or step of the challenge of translating nonverbal behavior. This was a deliberate choice which was meant to allow for the individual improvement of any individual step without prejudice to the overall strategy presented in Chapter 1. Clearly there are serious technological hurdles which need to be overcome before the translation of nonverbal behavior can become a reality.

6.2 Challenges and Future Developments

There is no doubt possible that international companies should encourage the development of well designed Nonverbal Behavior Cultural Translators. These tools are the natural expansion of today's set of CSCW tools. As explained in Chapter 1, the increased distribution of teams and the want for effective collaboration across cultural boundaries are important obstacles in the development of international companies. It would therefore be to the advantage of international companies to promote the development of products like the NBCT. It is hoped that the present document will serve as a guiding reference for those seeking to make translated nonverbal behavior available in distributed environments.

While it is obvious that future development should start with the design and implementation of the NBCT, it is possible to envision another potential use of such a

product which has not been mentioned up until now. In particular, the NBCT could serve as a training tool to help international travelers become more familiar with the nonverbal mores of their future business partners. For example, an American business man could interact with a Japanese computer agent in order to practice his Japanese communication skills before going to Japan. This could be done using video input from the American (to follow the example), and is relatively easy to implement in a computer. The challenge lies in building a sufficiently sophisticated Japanese agent which will react realistically to the nonverbal behavior of the American. Because the practice session would not be scripted, this is a challenging problem to solve. However, such cross-cultural communication practice sessions could help alleviate some of the problems which people regularly encounter in face to face communications.

Despite its potential usefulness, the NBCT's use could have some negative effects, and these should be discussed along with the potential benefits. Because the NBCT will never be able to fully replace face to face contact in the quality of feedback provided, it would be dangerous if some people became overly reliant on such a product. In particular, too great a reliance on the NBCT might take away from a person's face to face communication skills, as he or she becomes more expert at interacting with people through sheltered computer supported interactions than in face to face settings. However, it must be remembered that similar arguments were probably made at the time of the telephone's first rise, and that the telephone has not replaced face to face contact or taken away from people skills at face to face communication.

In a related problem, the NBCT could make people lazy as they rely on computers to translate nonverbal behavior for them, thus avoiding to effectively learn about other cultures. This could have disastrous effects when individuals end up meeting face to face, without the support of the NBCT which they have grown reliant on. A similar phenomenon has been observed with respect to spell checkers. The prevailing mindset is that learning how to spell is not necessary since computers can automatically make up for human shortcomings in this area.

The final danger with the NBCT lies in that it provides an illusion of reality –the better the illusion, the better the NBCT– and that this illusion might become so convincing as to overcome the reality. However, it will be many years until the quality of capture, translation, and representation of human behavior by computers can match that of humans, and so the problem is not a real one at this stage.

Finally, as alluded to in Chapter 3, the future development of the NBCT hinges on collaboration between computer scientists, graphic artists and animators, and psychologists. This type of collaboration is still in its early phases, but as it becomes more common, the theoretical tools needed from psychologists for developing algorithms performing the characterization, interpretation, and translation of nonverbal behavior will become more readily available.

As outlined in this conclusion, translation of meanings between cultures and representations of nonverbal behavior are beset with several important issues. However,

simple methods can be used to provide "first order" solutions. These solutions could become more sophisticated as psychologists provide NBCT designers with more appropriate and definite tools. This thesis does not purport to provide a final solution, but only to provide some ideas and conclusions for future work, and for the eventual development of the NBCT.

Bibliography:

1. Adams, L., and Toomey, L., Designing a Trans-Pacific Virtual Space, ACM SIGGROUP Bulletin, v 19, n 3, 15-18, 1998
2. Bartlett, S. M., Hager, J., Ekman, P., Sejnowski, T., J., Measuring Facial Expressions by Computer Image Analysis, Psychophysiology, v 36, 253-263, 1999
3. Bavelas, J. B., Chovil, N., Coates, L., Roe, L. Gestures specialized for dialogue, Personality and Social Psychology Bulletin, V 21, n 4 (394-405) 1995
4. Beckmans, P., R., Behavioral Expression and Related Concepts, Behavior and Psychology, v 24, 85-97, 1996
5. Belot, L., Les Francais a l'épreuve du management japonais, Le Monde, April 14, 1999. http://www.lemonde.fr/actu/entreprise/auto/renault/990414/management.html
6. Birdwhistell, R., Kinesics and Context, University of Pennsylvania Press: Philadelphia. 1970
7. Blumberg, B., PhD Thesis, MIT Media Lab, 1997
8. Bourke, P., Textures, http://www.mhri.edu.au/~pdb/texture/, 1999
9. Boyer, D. G., Handel, M. J., and Herbsleb, J., Virtual Community Presence Awareness, ACM SIGGROUP Bulletin, v 19, n 3, 11-14, 1998
10. Carroll, C., Evidences Invisibles: Americains et Français au Quotidien, Seuil: Paris, 1987
11. Cohn, J. F., Zlochower, A. J., Lien, J., Kanade, T., Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding, Psychophysiology, v 36, 35-43, 1999
12. Dreyfus, H., What Computers Still Can't Do, MIT Press, Cambridge, MA, 1992
13. Efron, D., Gesture and Environment, New York, NY, King's Crown, 1941. Cited in Ekman, 1969
14. Ekman, P., and Friesen, W.V., The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding, Semiotics, v 1, 49-98, 1969
15. Ekman, P., Emotions in the Human Face Cambridge University Press: Cambridge (128-141) 1982
16. Ekman, P., Friesen W. V., Facial Action Coding System, Consulting Psychology Press, Palo Alto, CA, 1978
17. Fridlund, A. J., Human Facial Expression, Academic Press, Inc: San Diego, 1994

18. Friesen and Ekman, undated, cited by Oster et al. (1992)
19. Fuchs, L., Poltrock, S., Wojcik, R., Business Value of 3D Virtual Environments, ACM SIGGROUP Bulletin, v 19, n 3, 25-29, 1998
20. Gercik, P., On Track with the Japanese, Kondansha America: New York, N.Y., 1996
21. Goldin-Meadow, S., When Gestures and Words Speak Differently, Current Directions in Psychological Science, v 6, n 5, 138-143, 1997
22. Hiltz, S. R., and Turoff, M. , "The Network Nation, Human Interaction via Computer", Addison-Wesley Publishing Company, Reading, MA, 1978
23. Jensen, R., Animation Engineer, Pixar, Inc., Personal Communication, April 23, 1999. email: rjjensen@alumni.princeton.edu
24. Labarre, Weston, The cultural basis of Emotions and Gestures, Journal of Personality, v 16, 49-68, 1947
25. Lagardère, F., cited in "La creme de canard, 24 Mars 1999", web-based edition of Le Canard Enchainé, http://www.electriccafe.org/Canard/CC990324.html, March 24, 1999
26. Lewicki, R. J., Bunker, B. B., Developing and Maintaining Trust in Work Relationships, in Trust in Organizations, Frontiers of Theory and Research, Kramer, R., M., Tyler, T. R., editors, Sage Publications, London, 1996.
27. Mantovani, Giuseppe, New Communication Environments, from Everyday to Virtual, Bristol and Francis, Bristol, PA, 1996
28. Mark, Gloria. Building Virtual Teams: Perspectives on Communication, Flexibility, and Trust, ACM SIGGROUP Bulletin, v 19, n 3, 38-41, 1998
29. Marrin, T., Picard, R., Analysis of Affective Musical Expression With the Conductor's Jacket, Proceedings of the XII Colloquium on Musical Information, Gorizia, Italy, September, 1998.
30. McCarthy, J., Book Review, April 10th, 1996, verified May 11, 1999 http://www-formal.stanford.edu/jmc/reviews/dreyfus/node3.html
31. McGrath, Andrew, The Forum, ACM SIGGROUP Bulletin, v 19, n 3, 21-25, 1998
32. Microsoft Corporation, Microsoft Chat Home, http://www.microsoft.com/windows/ie/chat/, April 27, 1999
33. Mignault, A., Research Assistant, VISMOD Group at the MIT Media Lab. Personal Communication, April 23, 1999. email: facial@media.mit.edu
34. Oster, H., Hegley, D., and Nagel, L., Adult judgement and fine grained analysis of infant facial expressions: Testing the validity of a priori coding formulas, Developmental Psychology, 28, 1115-1131, 1992
35. Palisades Research, 3D Builder Pro General Information, http://www.findthem.com/release.htm, 1999
36. Picard, R. W., Affective Computing, MIT Media Lab Perceptual Computing Section Technical Reports No. 321, 1995
37. Picard, R. W., Cosier, G., Affective Intelligence –the missing link?, British Telecom Technology Journal, v 15, n 4, 1997
38. Picard, R., Affective Computing, MIT Press: Cambridge, MA. 1997.
39. Picard, R., Professor of Media Art and Sciences, MIT, Personal Communication, Feb 22, 1999. email: picard@media.mit.edu
40. Sahlins, M., Islands of History, Chicago, IL, University of Chicago Press, 1985

41. Scheirer, J., Fernandez, R., Picard, R. W., Expression Glasses: A Wearable Device for Facial Expression Recognition, MIT Media Lab Perceptual Computing Section Technical Reports No. 484. Submitted to CHI 1999.
42. Shweder, R.A., and Sullivan, M.A., Cultural Psychology: Who needs it?, Annual Review of Psychology, v 44, 497-523, 1993
43. Starner, T., Weaver, J., Pentland, A., Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video, available at: http://vismod.www.media.mit.edu/tech-reports/TR-466/main-tr466.html, to appear in PAMI, submitted 4/26/96, 1996
44. Streek, J., Gesture as Communication: its Coordination with Speech and Gaze, Communication Monographs, v 60, 275-299, 1993
45. Stroud, M., Emotions over a Wired, Solved?, Culture News from Wired News, http://www.wired.com/news/news/culture/story/19356.html, April 27, 1999
46. Thomas, F., Johnston, O., The Illusion of Life, Disney Animation, Hyperion: New York, 1981
47. Tyler, T., Kramer, R., Wither Trust, in Trust in Organizations, Frontiers of Theory and Research, Kramer, R., M., Tyler, T. R., editors, Sage Publications, London, 1996.
48. Uderzo, A., De Flamberge à Astérix, Paris : Philipsen, 1985
49. Vilhjálmsson, H., Research Assistant, Media Lab, MIT, Personal Communication, April 26, 1999. email: hannes@media.mit.edu
50. Vilhjálmsson, J. H., Cassel, J., BodyChat: Autonomous Communicative Behaviors in Avatars, in ACM Proceedings of the Second International Conference on Autonomous Agents, Minneapolis, May 9-13, (269-276), 1988
51. Walterson, G., Calvin and Hobbes Web site, cropped from cartoons published April 25, 1988, and March 29, 1988 http://www.calvinandhobbes.com/ieindex.html
52. Walther, J. B., Relational Aspects of Computer Mediated Communication: Experimental Observations over Time, Organizational Science, v 6, n 2, 186-203, 1995
53. Watzlawick, P., Guide Non-Conformiste à l'Usage de l'Amérique, Seuil: Paris, 1987
54. Webster's Third New International Dictionary of the English Language Unabridged, 1981
55. Williams, J., Research Scientist in the speech, vision, and robotic group (SV & R) at Cambridge University, England, Personal Communication, April 3, 1999. email: jdw30@eng.cam.ac.uk
56. Worlds Inc., Worlds Inc. Home Page, http://www.worlds.net and PR 005, http://www.worlds.net/news/PressReleases/prn005.html, April 27, 1999
57. Wren, C. R., Pentland, A. P., DYNAMAN: A Recursive Model of Human Motion, available at http://vismod.www.media.mit.edu/tech-reports/TR-451/index.html, to appear in: Image and Vision Computing, 1998
58. Yacoob, Y., Davis, L. S., IEEE Transactions on Pattern Analysis and Machine Intelligence, v 18, N 6, (636-642) 1996

2655-40