HUMAN TECHNOLOGY

An Interdisciplinary Journal on Humans in ICT Environments ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 2 (1), April 2006, 55-83

ASSISTIVE TECHNOLOGY AND AFFECTIVE MEDIATION

Nestor Garay
Computer Science Faculty
University of the Basque Country, Spain

Idoia Cearreta
Computer Science Faculty
University of the Basque Country, Spain

Juan Miguel López
Computer Science Faculty
University of the Basque Country, Spain

Inmaculada Fajardo
Computer Science Faculty
University of the Basque Country, Spain

Abstract: *The lack of attention towards affective communication in assistive technology and disability research can be potentially overcome thanks to the development of affective computing and affective mediation areas. This document describes the main impairments and disorders that may involve affective communication deficits. We also present several affective mediation technologies that are being applied or may be integrated in assistive technologies in order to improve affective communication for a range of disabilities. Finally, we describe our experience with Gestele, an application that incorporates multimodal elements of affective mediation for people with mobility impairments, as well as the results of an empirical study conducted to evaluate the application's validity in communication via telephone.*

Keywords: *people with special needs, assistive technology, affective communication deficits, affective mediation.*

INTRODUCTION

The abundance of research in the assistive technology area is making it possible for people with special needs, mainly disabled people, to communicate, work, and perform daily activities that they were not able to perform efficiently and autonomously before the existence of these technological aids. Within assistive technology, the Augmentative and Alternative Communication (AAC) research field is used to describe additional ways of helping people who find it hard to communicate by speech or writing. AAC helps them to communicate more easily (International Society for Augmentative and Alternative Communication [ISAAC], n.d.).

DOI: http://dx.doi.org/10.17011/ht/urn.2006159

AAC includes many different methods. Signing and gestures that do not need any extra devices are called unaided systems. Other methods use picture charts, books, and special computers. These are called aided systems. AAC can help people understand what is said to them as well as help them to say and write what they want (ISAAC, n.d.). However, in the specific area of communication, there is an aspect that has received little attention from assistive technology researchers: affective communication.

Human beings are eminently emotional, as their social interaction is based on the ability to communicate their emotions and to perceive the emotional states of others (Casacuberta, 2001). However, a wide range of disabilities involve deficits in the different stages of affective processing (sensing, expressing, or interpreting affect-relevant signals). Consequently, people with these kinds of disabilities can be considered emotionally handicapped (Gershenfeld, 2000). *Affective computing*, a discipline that develops devices for detecting and responding to users' emotions, and *affective mediation*, computer-based technology that enables the communication between two or more people displaying their emotional states (Garay, Abascal, & Gardeazabal, 2002; Picard, 1997), are growing research areas (Tao & Tan, 2005) that must join assistive technology research to improve the neglected area of affective communication in disabled people.

The Laboratory of Human-Computer Interaction for Special Needs (LHCISN) research group at the University of Basque Country, Spain, is currently devoting efforts to such an aim. These efforts are materialized in the *Gestele* prototype, a multimodal and multistage affective mediation system for people with mobility and speech impairments (Garay-Vitoria, Abascal, & Gardeazabal, 2001). In the following pages, we present the definition of assistive and affective mediation technologies. Second, we describe the main affective impairments or disorders that create communication difficulties. Third, we present a review of the effort devoted to improving affective communication in both the assistive technology and the affective computing areas. Finally, we focus on the description of the Gestele system and an empirical study conducted to validate its affective communication.

ASSISTIVE TECHNOLOGY

Assistive technology is defined by King's Fund Consultation as "any product or service designed to enable independence for disabled and elderly people" (Foundation for Assistive Technology, n.d.,¶1). This technology is making great progress due to several reasons, such as technological advances, legislation, research and development programs (both within Spain and internationally), the proliferation and relevance of users' associations, and so forth. This field includes knowledge from sources such as computer science, telematics, and robotics, and, more specifically, fields such as human-computer interaction (HCI), domotics, artificial intelligence, multimedia technology, ergonomics, to name a few. Assistive technology has to cope with a variety of problems, such as detecting users' needs, evaluating results, social and ethical issues, the issue of affordability, and the use of a technology appropriately related to the problems at hand. It is interesting to note that the process of making devices accessible to people with special needs frequently causes such devices to become more accessible for the entire population (Vanderheiden, 1998).

Assistive technology research has made great advances in the area of the verbal, or explicit, communication channel. The LHCISN especially has made a notable effort in the

development of *communicators*, that is to say, machines (with their corresponding software) that allow users with motor and oral disabilities to communicate (Garay-Vitoria, 2001; Gardeazabal, 2000). Communication usually takes place through the generation of messages in various forms to be received by the interlocutors. This technology usually serves to write messages that can be seen on a screen device, be printed, or be synthesized via a text-to-speech system. The major areas of research are presented here.

Reduced Keyboards

A keyboard is called a reduced keyboard when the number of keys, "k," is fewer than the number of selectable characters, "c," which is the cardinal number of the selection set (Gardeazabal, 2000). These keyboards prove to be extremely useful, for example, when sending SMS (Short Message Service) communications on mobile phones, as well as helping users with special needs. The usual key distribution is the T9 keyboard[1], however, there are studies that try to obtain better distributions (Arnott & Javed, 1992; Foulds, Soede, & Van Balkom, 1987; Gardeazabal, 2000; Lesher, Moulton, & Higginbotham, 1998a; Levine, Goodenough-Trepagnier, Getschow, & Minneman, 1987). The disambiguation that the reduced keyboards achieve can also be used as a prediction system.

Scanning Set Distribution

In scanning input, the distribution of the selection set is crucial to optimize the time required to compose messages. A number of studies in the literature address this matter (Gardeazabal, 2000; Lesher, Moulton, & Higginbotham, 1998b; Venkatagiri, 1999). For example, Gardeazabal (2000) studied the optimal distribution with bidimensional and tri-dimensional matrices. It was observed that the access time changes when taking item frequencies into account. In two dimensions, square matrices are a valid choice. With more than two dimensions, several distributions that achieve very good results are found, and usability studies have to be carried out in order to determine the best for each particular user.

Automatic Adaptation of the Scanning Period

Scanning-and-selection input systems scan the selection set with a fixed time period "T". Text production is highly influenced by this parameter. Large "T" values produce longer text composition times. Therefore, it is important to maintain "T" values as short as possible. However, it has been observed that too short a "T" value increases the number of mistakes made by the user, and therefore affects the time of message composition. Gardeazabal (2000) presents systems based on fuzzy logic and traditional logic, taking into account factors such as fatigue, error rate, and so on, to smoothly adapt the scanning period to specific users.

Text Prediction

Prediction techniques use contextual information to anticipate what a person is going to write (Garay & Abascal, 1994). If the system is able to guess correctly, the number of keystrokes needed to write a sentence decreases. Apart from enhancing communication speed, the physical effort required to compose messages is reduced. In addition, the prediction software

may also fix spelling mistakes, reorder sentences and, in general, enhance the quality of the composed messages (Magnuson, 1995; Zordell, 1990). Prediction systems can be seen as intelligent agents that assist users in composing their texts. They capture user inputs to make guesses and provide a list of items as outputs. These outputs can be incorporated into the applications used to compose texts by trying to emulate user behavior. Hence, the most advanced predictors have learning features that are able to make inferences, are adaptable, and act independently. In certain cases, these technologies may converse with users, mainly making personal vocabulary adaptations. These predictors may benefit people with severe motor and oral disabilities, such as cerebral palsy or hemiplegia.

AFFECTIVE MEDIATION

As the brief review in the previous section shows, a great effort has been carried out in order to improve verbal communication. However, the emotional features of verbal communication and the improvement of the implicit, or nonverbal, communication traditionally have received less attention. Therefore, in focusing on the improvement of the affective communication for disabled people, assistive technology research must take the research made in the field of affective mediation into account.

As Figure 1 shows, the main objective of affective computation is to capture and process affective information with the aim of enhancing the communication between the human and the computer. Within affective computing, affective mediation uses a computer-based system as an intermediary among the communication of certain people, reflecting the mood the interlocutors may have (Picard, 1997). Affective mediation tries to minimize the filtering of affective information of communication devices; filtering results from the fact that communication devices are usually devoted to transmission of verbal information and miss nonverbal information. Affective mediation has a direct application within AAC. There are also other applications in this type of mediated communication, for example, textual telecommunication (affective electronic mail, affective chats, etc.). As previously mentioned, applications developed in AAC area are useful for both disabled and nondisabled people.

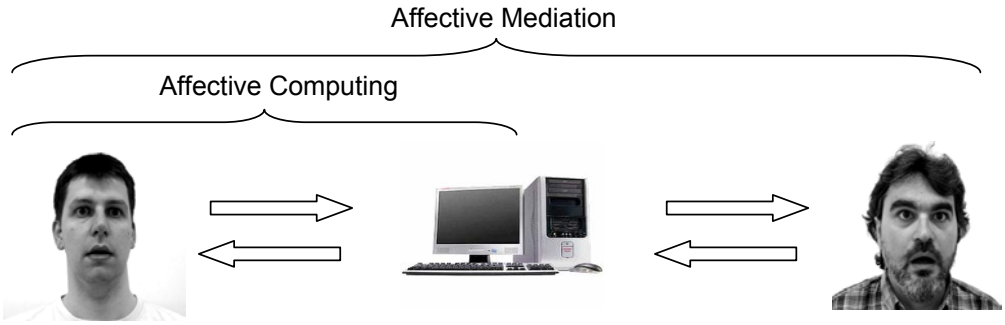

Figure 1. Affective Computing vs. Affective Mediation

IMPAIRMENTS AND DISORDERS INVOLVING AFFECTIVE COMMUNICATION DEFICITS

With the aim of organizing affective computing research, Hudlicka (2003) proposes a classification based on different stages of affect processing, such as sensing, recognizing, interpreting, selecting, and expressing affects. In this document, we use this classification to organize those impairments involving affective communication disabilities. Table 1 shows different types of impairments or disorders signaling the pertinent stage of affective communication.

Table 1. Impairments or Disorders Involving Deficiencies in Affective Processing and the Affective Computing Technologies Useful in Addressing the Deficiencies.

Impairment/Disorder	Impaired stage of affective communication	Useful affective computing technology	Application contexts
Visual impairments (low vision, blindness)	▪ Sensing visual affective information: face and body gestures	▪ Facial affect recognizer ▪ Emotional text readers	▪ Chat or videoconferences
Hearing impairments (pre- & postlocative deafness, etc.)	▪ Sensing speech prosody (pitch, volume or velocity) ▪ Expressing speech prosody	▪ Prosody/speech affect recognizer ▪ Emotional speech synthesizer ▪ Text affective recognizer	▪ Telephone communication ▪ Chat
Learning disorders (e.g., dyslexia)	▪ Interpreting emotions in facial expression (exclusive of a visual perceptual subtype of dyslexia called Irlen syndrome)	▪ Facial affect recognizer ▪ Emotional avatars ▪ Text affect recognizer	▪ Chat ▪ Diagnosis ▪ Training ▪ Biofeedback
Mobility impairments (locked-in syndrome, apraxia, etc.)	▪ Expressing nonverbal information (postural and facial gestures, speech prosody)	▪ Prosody/speech affect recognizer ▪ Emotional speech synthesizer ▪ Sensing devices for psycho- physiological affect recognition (EEG, EKG, skin conductance, eye tracking)	▪ Human-to-human communication ▪ Telephone communication ▪ Proactive/predictive interfaces (Moshkina & Moore, 2001)
Developmental disorders (autism, Asperger syndrome, etc.)	▪ Use of multiple nonverbal behaviors (ocular contact, facial expressions, corporal gestures, etc.) ▪ Recognition of emotions in face and speech	▪ Facial affect recognizer ▪ Emotional Avatars ▪ Prosody/speech affect recognizer ▪ Emotional speech synthesizer ▪ Sensing devices for psycho-physiological affect recognition (EEG, EKG, skin conductance, eye tracking) ▪ Emotional Hearing Aids (Birkby, 2004)	▪ Human-to-human communication ▪ Diagnosis ▪ Training ▪ Treatment ▪ Biofeedback
Anxiety disorders (social or specific phobia)	▪ Affect interpretation and expression: irrational fear to specific stimuli or situations	▪ Sensing devices for biofeedback (EEG, EKG, skin conductance, eye tracking) ▪ Virtual reality	▪ Treatment (systematic desensitization)

Sensing External or Internal Affect-Relevant Signals

Emotional information is transmitted mainly through two channels of human communication: the verbal or explicit channel (language) and the nonverbal or implicit channel (Cowie et al., 2001; Knapp, 1980). Therefore, the affect-relevant signals at the sensing stage, for both individuals and machines, can be verbal (speech or text) and nonverbal (facial expression or voice intonation). Among the impairments that would present problems in this stage of affective processing, we find the category of sensorial deficits, mainly visual and hearing impairments (Jacko, Vitense, & Scott, 2003).

Visual impairments are divided into two general categories: blindness and poor vision. Individuals with blindness have absolutely no sight, or have so little that learning must take place through other senses. Poor vision includes dimness, haziness, extreme far-sightedness or near-sightedness, color blindness, tunnel vision, and so forth. Hearing impairment is defined as a lack or reduction in the ability to hear clearly due to a problem somewhere in the hearing mechanism (Jacko et al., 2003). It is interesting to mention that most deaf people learn a different modality of language, the sign language, which has prosody to express affects like in an oral language. For unmediated communication, it is evident that a blind person cannot sense visual affect-relevant signals such as facial expressions or postural gesture of the interlocutor, while deaf people process neither the explicit emotional message transmitted by the speech nor the voice intonation associated with it.

According to E. T. Hall (1998, p. 53), "although we tend to regard language as the main channel of communication, there is general agreement among experts in semiotics that anywhere from 80 to 90 percent of the information we receive is not only communicated nonverbally but occurs outside our awareness." Meanwhile, Mehrabian (1971) affirms that 7% of a message between two people is verbal, 38% vocal (tone, shades, etc.) and 55% is body language. Therefore, in terms of Mehrabian, unaided visually impaired and deaf people lose 55% and 38%, respectively, of the affective information that people without those sensorial impairments are able to process.

Recognizing an Affective State

Several developmental disorders cause problems with the ability to recognize, to comprehend, the affective states of other people. In this category, we include developmental disorders such as autism or Asperger syndrome, and learning disorders such as the so-called visual dyslexia.

Regarding autism, it is interesting to highlight that one of its diagnostic criteria is the deficit in the use of nonverbal social communicative behaviors (e.g., eye-to-eye gaze, facial expression, or social reciprocity). One noted problem of these individuals is their great deficit in the recognition of faces (Boucher & Lewis, 1992; Klin et al., 1999; Tantam, Monaghan, Nicholson, & Stirling, 1989) and facial expressions, in particular (Capps, Yirmiya, & Sigman, 1992; Celani, Battacchi, & Arcidiacono, 1999; Pelpherey et al., 2002). For instance, Pelpherey et al. (2002) asked a group of autistic and a group of nonautistic adults to identify the emotion portrayed in a set of pictures of facial expressions from the database of Ekman and Friesen (1976). Researchers found the autistic group identified a smaller percentage of the emotions in comparison to the nonautistic group (concretely, fear was the emotion that established the difference). Furthermore, the analysis of the visual scan path revealed that

individuals with autism scanned relevant areas of the faces (e.g., eyes, nose, and mouth) less frequently than the control group.

In addition to the facial recognition deficit, autistic individuals can also possess a deficit in emotional prosodic recognition. This fact has been documented (e.g., VanLancker, Cornelius, & Kreiman, 1989) and, at the moment, the use of prosodic information is being considered for both diagnosis and enhancement of emotion recognition in autism disorders (Hall, Szechtman, & Nahmias, 2003). As we will see in the following sections, this data can be taken into account by researchers when designing affective mediation technology for people with autism.

On the other hand, the deficit in facial expressions recognition does not seem to be exclusive to disorders in the autism spectrum. In fact, a kind of dyslexia, named visual dyslexia or Irlen syndrome, also displays deficits in this area. Dyslexia is a learning disorder characterized by difficulties in single word decoding that is not a result of a generalized developmental disability or sensory impairment. Irlen syndrome is claimed to present a deficit in the magnocellullar visual neurological pathway (central nervous system) that would cause visual processing problems. Robinson and Whiting (2003) carried out a study contrasting a group of children with Irlen syndrome with a group of children with no learning disability. The participants performed a facial affect recognition task built with facial affect pictures from the Ekman and Friesen (1976) database. Researchers found that children with Irlen syndrome obtained lower recognition scores than the control group. The authors interpreted the findings as supporting the claims that individuals with Irlen syndrome are more likely to possess social interaction problems due to the deficits in the interpretation of subtle visual clues such as emotional facial expressions. Again, this problematic aspect of affective processing should be taken into account in the design of assistive technology for people with this kind of disorders.

Interpreting and Appraising the Current Situation to Derive, Select, or Generate an Affective State

Developmental disorders, such as autism or Asperger syndrome, or anxiety disorders, such as social or specific phobias, often result in an individual's inability to accurately assess, interpret, or generate an appropriate affective state. As mentioned before, people with autism spectrum disorders show qualitative alterations in their nonverbal communication abilities. Apart from the apparently proven deficit in the recognition of prosody and facial expressions, they possess a deficit in the interpretation of affective signals. For instance, an Asperger individual may recognize a happy facial expression, but not interpret what it means and how to respond to it.

An added problem for people with these kinds of disabilities is that they cannot judge other people's motivations. Hence, such individuals can be victimized by frauds or even worse exploitation. It may be useful to explore whether affect detection technology could help preventing these situations.

On the other hand, anxiety disorders such as social phobia are characterized by irrational fear toward the presence of an object or to a specific situation (spiders, flights, or social situations). Irrational fears can be interpreted as cognitive distortions that are manifested in an exaggerated and irrational emotional response. Affective computing technologies may be applied to assist phobic individuals in the treatment of these handicapping emotional responses.

Expressing the Affective State via One Established Behavioral Script

According to the bio-informational model proposed by Lang (1979, 1984), the emotional response can be produced through three systems: the subjective, or verbal; the behavioral; and the physiological. Impairments or disorders that compromise the correct operation of any of these systems can potentially affect the expression of affective states. In this category, it is possible to mention once again, sensorial impairments such as prelocutive deafness and mobility impairments such as locked-in syndrome or apraxia. People with prelocutive deafness have difficulties achieving functional levels of the oral and written language (e.g., Alegría, 1999; Asensio, 1989; Fajardo, 2005; Leybaert, Alegria, & Morais, 1982; Silvestre & Valero, 1995), so emotional communication through these ways is deficient. Regarding mobility impairments, one of the most relevant examples is the locked-in syndrome, which is characterized by a complete paralysis. In these individuals, verbal self-reporting is difficult and slow because they frequently have to use direct brain-machine interface technologies (Moshkina & Moore, 2001) or predictive keyboards (Gardeazabal, 2000) to communicate. Moreover, due to various reasons (such as problems in the articulation of facial gestures), people with apraxia also possess difficulties in the production of oral language, and consequently in the transmission of the verbal and nonverbal emotional aspects associated with these media (message and prosody). As we will see later, both types of mobility impairments could benefit from an assistive technology that helps expressing affective states.

So far, we have presented a partial list of impairments or disorders that may involve affective communication problems. We have tried to provide an example of the practical relevance of the interaction between affective mediation and assistive technology research areas. Next, we continue with the description of affective mediation technologies that can be used by disabled people.

APPLICATIONS OF AFFECTIVE MEDIATION FOR ASSISTIVE TECHNOLOGY

According to Hudlicka (2003), the challenges of affect in HCI encompass recognizing user affect, adapting to the user's affective state, generating affective behavior by the machine, and modeling the user's affective states or generating affective states within an agent's cognitive architecture. In this section, we present a number of the most relevant techniques to enhance affective communication for disabled people in the areas of affect recognition and generation.

Affect Recognition Technologies

As mentioned earlier, Lang (1979) proposed three systems involved in the expressions of emotions:

- Subjective or verbal information reports on emotions perceived and described by users;
- Behavioral information involves facial and postural expressions, speech paralinguistic parameters, and so forth; and
- Psychophysiological responses are expressed by heart rate, galvanic skin response (GSR), and electroencephalographic (EEC) response.

Consequently, with the aim of sensing and recognizing emotions, a computer must incorporate sensors to receive the emotional output from each of the expressive systems of an individual, and emotional models that allow for interpreting such signals. As Picard (1997) suggests, one of the main aspects of an intelligent computer is its ability to recognize emotions and infer an emotional state by looking at emotional expressions, and the capacity for reasoning a situation. In the same way, a recognizer might be a multimodal system, as it may have the ability to see and hear in order to make out facial expressions, gestures, and voice tone, as well as to recognize emotions from a text typed using a keyboard or spoken by the user. Therefore, the recognizer, apart from analyzing voice tone, gestures, and facial expressions, may also analyze the semantics of the words spoken or written by the user.

Some signals communicate emotions better than others, depending on the emotion, the person who is communicating them (with his/her possible disabilities), and the circumstances generating these emotions. Next, we present affect recognizer technologies that can be useful for various disabled people.

Facial Affect Recognizers

Facial affect recognizers try to obtain emotional information from facial expressions by analyzing the spatial configuration of different parts of the face, mainly the eyes, lips, and nose. It may be useful for those people who cannot process such information due to sensorial or visual difficulties or due to developmental disorders, such as autism or Asperger syndrome.

A facial affect analyzer may be helpful for autistic or Asperger individuals to understand the interlocutor's emotion because they have problems recognizing facial expressions. Birkby (2004) and Kaliouby and Robinson (2005) propose the Emotional Hearing Aid, which is a wearable device receiving input from a digital camera and uses the input file to interpret the mental and emotional state of the interlocutor and to inform the disabled user. As in the case of autism spectrum disorders, the Emotional Hearing Aid may be a suitable solution to assist children with Irlen syndrome to identify facial affects.

Unlike people with an autism spectrum disorder, the hearing and visual impairments that impact affective communication in some people are due not to a perceptive problem or comprehension problem but to a sensorial one. This is a relevant distinction for the design of affective mediation technology because the consequences are not identical. For example, deaf people, once they know that the interlocutor is sad, can produce the correct response (e.g., to comfort the other person). However, in the case of autism, even if the individuals know their interlocutor's emotion, they do not know how to react to it. For this reason, several technologies, such as the Emotional Hearing Aid, also incorporate a module that sends a recommended reaction to the user.

In order to identify a particular facial expression, individuals rely on the spatial configuration of the major features of the face, that is, the eyes, nose, and mouth (Diamond & Carey, 1986; Farah, Wilson, Drain, & Tanaka, 1998). Autistic individuals present a different pattern of face scanning, which some authors identify as underlying their face affect recognition deficits (Pelpherey et al., 2002). Therefore, sensing devices, such as the eye tracking devices, may be useful as both diagnostic and biofeedback methods for autistic individuals.

Visual impairment is another potential beneficiary from the facial affect analyzers. Although visually impaired people can use other nonverbal and verbal affective cues during the human-to-human interaction, there are contexts where such other cues are not available

and the visual ones must take primary relevance. Chats or videoconferences where the speech prosody may be interfered with or deficient are examples of such situations.

Speech Affect Recognizers

Prosody is a speech feature which, together with the linguistic content, conveys the emotion of the interlocutor. Therefore, the speech affect analyzers use voice parameters such as the pitch, volume, or speed of the speech (prosody elements) to indirectly infer the emotion of the speaker. There are diverse laboratories working on this kind of technology; for instance, the LHCISN is currently developing a prototype of a prosody affect recognizer. More specifically, for this type of disorder, the Oregon Health and Science University (OHSU, Portland, USA) laboratory is working on a project that will use speech and language technologies to improve the diagnosis, evaluation, and treatment of communication disorders, such as autism (OHSU, 2005). In the case of evaluation, that project will use speech technologies to measure the prosodic abilities in children's speech or to create speech stimuli to measure the children's abilities in understanding prosodic cues (this second case is addressed again when we discuss generating technologies).

In contrast to visually impaired, visual dyslexic, or autistic individuals, deaf people seem to have a special ability to process facial expression, mainly because they are better than their hearing peers in recognizing faces (Arnold & Mills, 2001). Deaf people may develop this ability in order to obtain emotional information from faces, since the prosody of speech is out of their scope. The lack of prosodic information would not be a problem in human-to-human communication where there is another source of information (nonverbal), but may be relevant in the use of text-telephones, a technology widely used by deaf people. In such cases, the speech prosody recognizers may prove to be useful for deaf people when trying to understand the affective state of the speaker. Accenture Technology Labs are developing an audio analysis technology for emotion recognition to help call center personnel to recognize the emotion of callers (Petrushin, 2000). Although the Accenture product was not intended for deaf people, it may be easily adapted and used by this kind of user.

One obvious and useful application for speech recognition technology is to provide and enhance human-to-human communication for disabled individuals. For hearing-impaired individuals, the interlocutor's words may be recognized and subsequently produced in the form of corresponding animated sign language, which appears on a display screen (Noyes & Frankish, 1992).

Sensing Devices

These devices are sets of psychophysiological response sensors that measure factors such as skin conductance, heart rate, pupil size, and so on. Applications using sensing devices must be able to relate the psychophysiological signals to the emotional responses of the user. Sensing devices are able to collect data in a continuous way without interrupting the person. The Affective Computing research group at Massachusetts Institute of Technology (MIT) in the USA (see Affective Computing, n.d.) is interested in many different sensing devices useful in recognizing affective signals. This research group has developed a prototype of a physiological detection system (Prototype Physiological Sensing System). This is a light and portable system, and relatively robust regarding changes in the user's position. The prototype

is based on a combination of four different sensors: the blood volume pulse (BVP) sensor, the galvanic skin response (GSR) sensor, the electromyogram (EMG) sensor, and the respiration sensor. The output from the prototype is ported to a computer for processing and recognition.

Such recognition technology may be useful for people with mobility impairments, developmental disorders (autism or Asperger), or hearing impairments. For instance, Moshkina and Moore (2001) propose that since locked-in people have continuous life monitoring systems, it would be possible to easily capture psychophysiological responses related to their emotional state. In some cases, where traditional measuring systems are too invasive, a kind of wearable computer exists to collect physiological signals[2]. Lisetti, Nasoz, Lerouge, Ozyer, and Alvarez (2003) are working on mapping these signals to emotional states, and they currently use this application for telehome health care. These kinds of applications may prove to be useful for people who are in need of a continuous assistance.

As in the case of face and speech affect recognizers, psychophysiological technologies can be added to the telephones used by hearing-impaired people. This way, these technologies may obtain from or provide to the interlocutor the emotional information that cannot be extracted or provided by means of other sources (pitch or volume of speech). In the same way, since people with autism are not able to recognize the emotional states of the interlocutor well, sensing devices, facial emotional information, and speech recognizers may serve as artificial recognizers to teach them this skill progressively.

Linguistic Affective Analyzer (LAA)

There are two primary types of LAAs: semantic systems and statistical systems. Whatever the underlying system, the objective of this technology is to abstract the emotional information from the linguistic content of text or speech. The potential users of the LAA would be people with reading or language difficulties, such as prelocutive deaf people and people with autism or dyslexia. Several studies have been performed in order to obtain an affective value for written text (e.g., Bradley & Lang, 1999; Strapparava & Valitutti, 2004). For example, the Affective Norms for English Words (ANEW) provide a set of normative emotional ratings for a large number of words in the English language (Bradley & Lang, 1999). Each word has three associated values that correspond to three emotional dimensions: arousal, valence, and control (according to the bio-informational model proposed by Lang, 1995).

Affect Generating Technologies

Within the category of affect generating technologies, we include all the technologies that display the emotional information transmitted by the user, recognized by any of the previous technologies or automatically generated by the computer. Emotion recognition is often considered as a part of emotion analysis. Synthesis is the opposite of analysis; it is the construction of the emotion. Both recognition and synthesis are able to work on the same system: the recognition can proceed to synthesize or generate several options, asking which one seems to be closer to what is perceived by people. This approach of the recognition is sometimes called “synthesis analysis” (see Affective Computing, n.d.). This section presents some of the most relevant technologies related to affect generation.

Emotional Speech Synthesizer

A speech synthesizer is able to convey the emotional state of the users by means of variations in voice parameters, such as pitch, speed, or volume. This can make the synthesizer provide its messages more satisfactorily. Unfortunately, it is difficult to limit the minimum number of necessary parameters to appropriately generate each basic emotion (just as it is hard to delimit the basic universal emotions). Authors such as Cowie et al. (2001) and Schröder (2003) provide a list of other researchers who suggest their own parameters and basic emotions. Several of the most recent tendencies in voice synthesis are based on the concatenation of previously recorded units. Concatenative synthesis is carried out by joining digitalized phoneme and syllable recordings in a coherent way; these recordings are recorded by a speaker and are stored in a database. Voice synthesizers are often called TTS (text-to-speech), as the voice synthesis is the process of creating speech from text.

Although visually disabled people have a normal development of affective communication abilities, they usually must use assistive technologies, such as screen readers with voice output. This output may be emotionally plain and unpleasant. In these cases, the emotional speech synthesizer could make the use of these devices more satisfactory. The same logic can be applied to the cases of hearing- or motor-impaired people. Those deaf people whose speech is difficult to understand usually make use of TTS technology, for instance, to maintain telephonic conversations (Garay et al., 2002; Garay-Vitoria et al., 2001). Motor-impaired people use speech synthesizers not only in telephone communication contexts, but also during human-to-human communication. For this reason, this technology may be one of the more beneficial for this group.

Affective Avatars

Avatar is a Sanskrit word that means the incarnation of a god on earth, and this term has been adopted into computer science via the gaming and 3-D chat worlds. The avatar is the visual "handle" or the display appearance a person uses to represent herself or himself. Thanks to graphical visualization developed for gaming purposes, emotional information can be displayed visually in an easy and comprehensible way. Avatars are actually more centered in physiological information, but behavioral information can be included as well. Three-dimensional and real-time graphics are used to represent a user's signals, and the goal is to provide unique, innovative, and discrete ways to recover data.

The inclusion of affect elements in avatars is being developed by multiple researchers (e.g., De Rosis, Pelachaud, Poggi, Carofiglio, & De Carolis, 2003; Lisetti et al., 2003), and have in mind several objectives (e.g., telehome health care, teleassistance, entertainment, computer as companion). Persons with almost any type of disorder or impairment potentially can benefit from this technology. Anthropomorphic affective avatars, such as the one proposed by Lisetti et al. (2003), can serve as a virtual tutor that teaches autistic people how to recognize facial expressions. Another application found for these systems was the possibility of adapting to the user's states dynamically, which would be very useful for mobility-impaired people. For example, De Rosis et al. (2003) have developed Greta, a realistic 3-D embodied agent that may be animated in real time and is able to express verbal and nonverbal affective signals (mainly facial expressions).

Emotional Robots

Emotional robots may be used in a teaching context as well as avatars, and can serve as companions for elderly people. For example, the Kismet robot "is capable of generating a continuous range of expressions of various intensities by blending the basis facial postures" (Breazeal, 2003, p. 143). One important point about Kismet is that its efficiency has been tested with real users.

Virtual Reality

Virtual reality can simulate interactive worlds that can be experienced visually in three dimensions and provide feedback in multiple modes (auditory, tactile, etc.). It has been used for the treatment of phobic disorders by "immersing the patient in a synthetic universe in which the anxious stimulus is introduced in an extremely gradual and controlled way" (Roy, 2003, p. 179).

After this review, it is clear that there is a great diversity of affective computing technologies that are being or can be used by people with disabilities. However, one important deficiency of these technologies is that they are usually unimodal and unistage, that is, they usually allow for the assistance in only one stage of the affective processing (e.g., sensing) and in one exclusive modality (e.g., visual). We present our own experience in the field by describing Gestele, a specific assistive technology designed to provide affective communication for oral- and motor-impaired people at diverse stages (modeling and expressing) and modalities (speech and text).

GESTELE: A MULTIMODAL AFFECTIVE MEDIATION TECHNOLOGY

Speech is the most common form of communication among people. According to Alm, Arnott, & Newell (1992), it has been estimated that the achieved speed in a normal conversation is about 150-200 words/minute, mainly in the case of the English language. Yet people with certain oral or mobility impairments are not able to use this effective communication channel. Therefore, Gestele was developed with the aim of assisting people with severe motor and oral communication disabilities. It is a mediation system that includes emotional expressivity hints through voice intonation and preconfigured small talk. The Gestele system is used for telephone communication (see Garay et al., 2002; Garay-Vitoria et al., 2001; Garay-Vitoria, Abascal, & Urigoitia-Bengoa, 1995). Gestele allows for different input options (directly or by scanning, depending on user characteristics) and makes holding telephone conversations involving affective characteristics possible for people with no speaking ability. This may be seen as a TTS system applied to telephone conversations by providing help in text message composition.

A telephone conversation can be modeled with a state-transition automaton, such as the one shown in Figure 2 (Garay-Vitoria et al., 1995). E0 is the state when the phone is not being used. In E3, sentences are precomposed for a future conversation. Prepared messages are classified according to their topic in order to allow the minimum effort possible for their

selection. The other states are the typical ones occurring during a telephone conversation. System initial interface is shown in Figure 3.

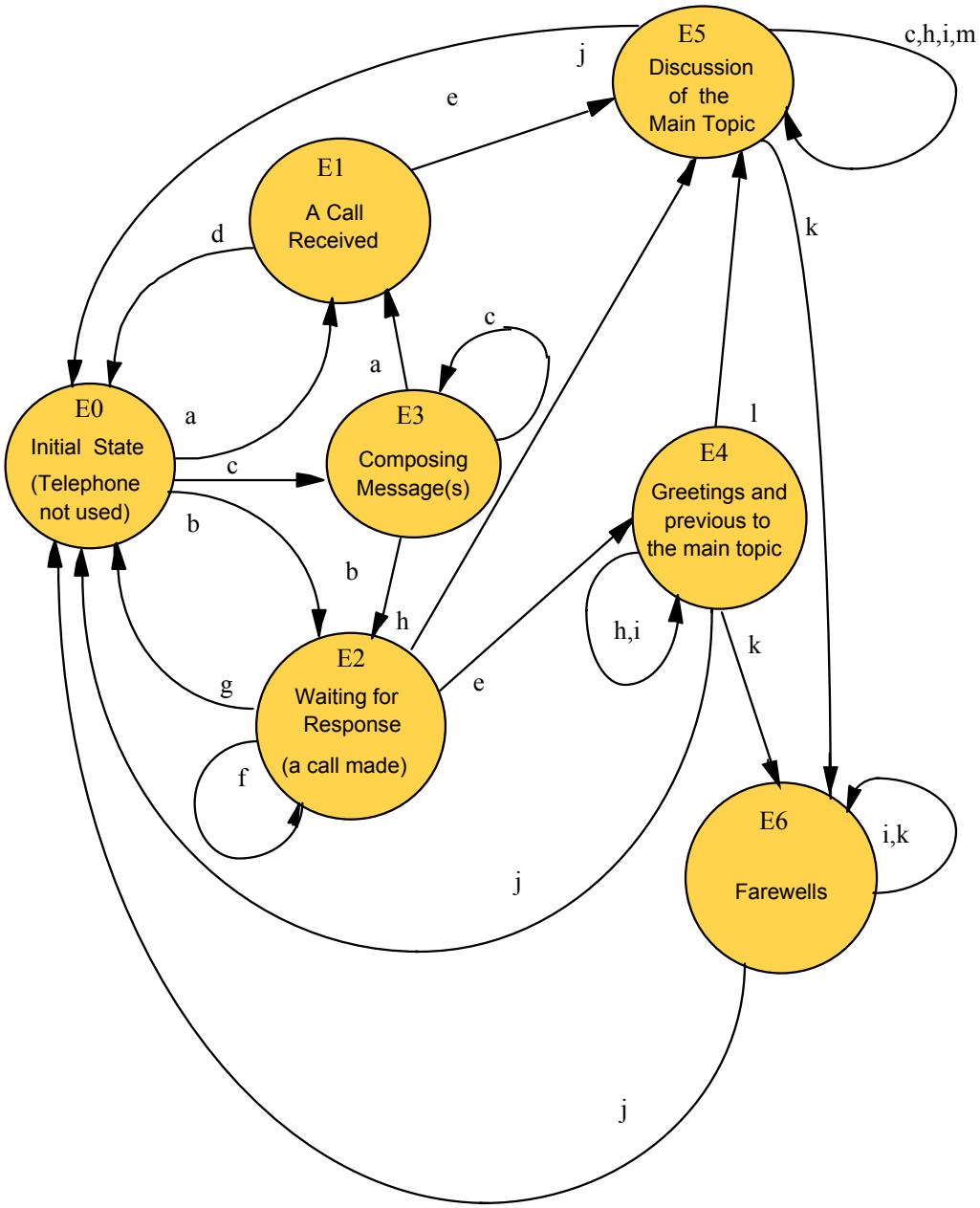

Figure 2. The automaton of the telephonic conversation aid. Notation for the arcs: a) receive a call; b) make a call; c) prepare a message; d) do not answer; e) salute/give a warning; f) asking for confirmation; g) there is no answer; h) send a message; i) send a filler remark; j) hang up the phone; k) say good-bye; l) introduce the main topic; m) give a warning for composing a message.

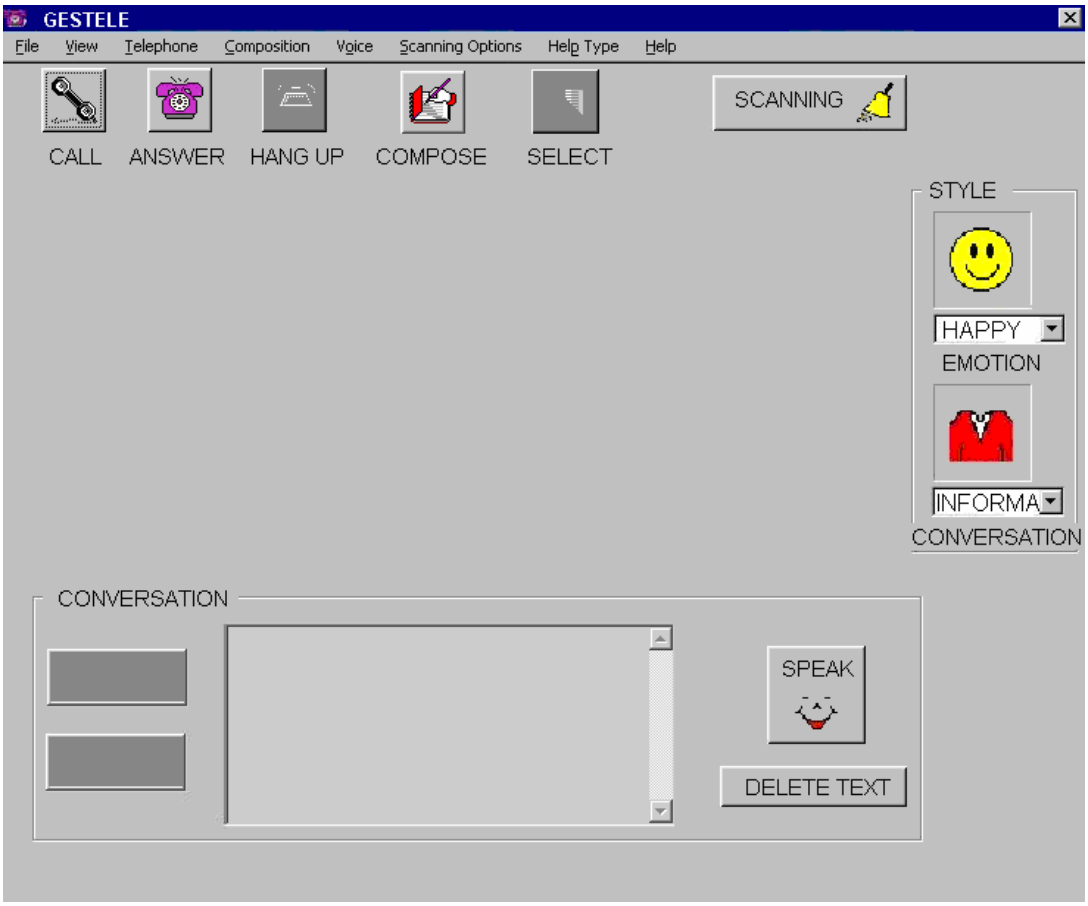

Figure 3. Gestele interface in E0 state of the automaton.

Advancing across this scheme, the user can produce meaningful sentences in the context of the conversation at a reasonable rate and effort. Therefore, in most states of the automaton, the previously mentioned problem of obtaining an acceptable message production speed is solved. When sentences have to be improvised, the user can obtain help from word prediction and composition accelerators, and syntactic and morphologic correctors to type his/her opinions in a reasonable rate of time and with minimal effort.

There are two factors to be kept in mind when building the model of conversation.

1. There is a set of social rules that guides conversational interaction. As conversation is a cultural matter, dialogue modeling is very dependent on the cultural context. Thus, a specific model might not be valid for people living in different countries, speaking different languages, and so on. For instance, the importance and duration of greetings, farewells, and polite questions, to name a few, are very different from one culture to another.
2. The conversation frequently affects the relationships between interlocutors. Consequently, the speaker can appear friendly and intelligent, and can manifest his/her points of view or prejudices. In the mediated dialogue, the interlocutor's personality should appear. If high user acceptation is desired, the model must able to automatically adapt itself to the user's personality.

Gestele allows sentence composition with a scan-based system that may be controlled by a single button. The distribution of the characters in the options matrix has been made by taking the frequencies of the letters in the Spanish language into account, as mentioned in Arruabarrena and González-Abascal (1988). The scanning speed is predefined but it can be changed using the options in the state bar of the interface. The options matrix is shown in the middle left part of the interface (see Figure 4).

To enhance communication speed, Gestele also has an associated word predictor that is activated while the user is composing messages. The prediction list is in the middle part of the interface (see Figure 4), between the controls relative to the selection of a number of pre-stored sentences (agreement, disagreement, etc.) and the conversational style. Selecting one of the options that are shown in the prediction list means that this option is included as the last word of the message composition, that is, in the text box in the middle-bottom of the interface.

With the aim of enhancing communication speed and minimizing silences during message composition, the system has pre-stored sentences made available automatically, depending on the state of the conversation. Once a sentence is written or selected, it can be then spoken via a commercial speech synthesis system connected to a modem with telephone line connection.

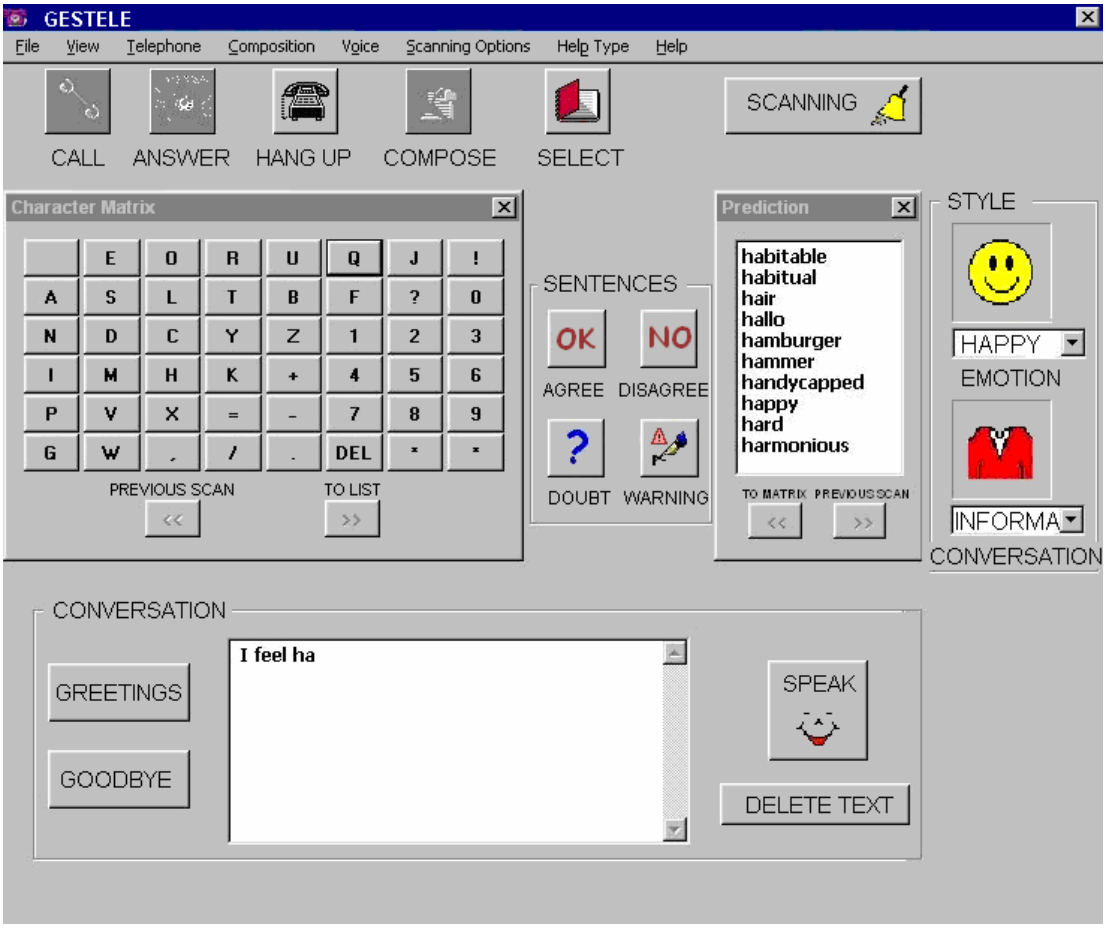

Figure 4. Gestele interface in E4 state of the automaton.

It has a conversation model similar to the one used in the CHAT system (Conversation Helped by Automatic Talk) by Alm et al. (1992) that allows the state of the conversation to be represented internally using an automaton (Garay-Vitoria et al., 1995, 2001). For instance, in an initial state, the system randomly selects a greeting among the stored ones. Then, introduction sentences can be produced, the main topic will be treated and, finally, the interaction concludes with a farewell. The system is also able to speak sentences stored, depending on the state of the conversation; the user controls the evolution of the dialogue at every moment by using the options presented in the interface (see Figure 4).

The main controls of the system are in the central area of the interface. Among them, there are buttons to

- call, answer and hang up
- compose and select a sentence
- speak via speech synthesis and stop the reproduction (delete text)
- generate filler remarks: agreement, disagreement, and doubts
- produce alert sentences to express that a new sentence is being composed, in order to avoid any silence being interpreted as a communication breakdown by the interlocutor
- say goodbye
- activate/deactivate the scanning option.

All these buttons cause effects on the automaton. They also may generate transitions in the automaton, depending on the current state and the required action.

The other controls are used to develop the conversation, but they do not change the state of the automaton. Sentences composed with the editor appear on the right side of the window. There are also two scrolling lists to change the speaker's emotional and conversation style in a simple way. Several utilities are associated with the editor, such as a virtual keyboard on the screen to write messages, either directly or using scan input. A word predictor window shows the most probable words starting with the letters the user has already written, in order to minimize the required keystrokes to compose messages (see Figure 4).

The state bar shows information relative to the controls and program execution. The menu shows all the possible program options. The majority of these options are more comfortable to use via the tool bar; however, several options have to be selected from the menu. The most usual are

- view, to select editor font
- voice options, relative to speech synthesis in order to change volume, pitch and speed
- scanning options, to configure scanning speed.

Sentences selected by the system are shown in the editor and synthesized via voice. It is also possible to voice the text in the editor window. The scanning option is given with the options on the main screen, changing the attention focus after a certain time.

At all times, the system must take the state of the conversation into account so as to offer speech performances with the most appropriate speech characteristics (mainly volume, pitch, and speed). These sentences have a particular purpose in the dialogue and are appropriate only in a given state. In this way, the achievable rate (measured in terms of words per minute) is enhanced and the needed keystrokes to compose sentences are minimized. Depending on the state of the conversation, the improvements achieved may vary. It is evident that in farewells and greetings, the improvements will be higher than in the main topic state. This results from the

main topic being particular for each dialogue, while farewells and greetings may serve for the entire set of dialogues. For this reason, as seen in Alm et al. (1992) and Garay-Vitoria et al. (1995), farewells and greetings are more predictable than main topic themes.

Another factor to bear in mind is the expression of personality and conversational style (Garay-Vitoria et al., 2001). The sentences should be adapted to the personality of the user (e.g., personalizing the user's speech). There is a need to study the user's vocabulary and the way he/she combines the words to create sentences with determined structures. The style of conversation is a parameter that depends on the user's frame of mind at a given time, as well as on the interlocutor he/she is addressing. With strangers, the adopted tone is normally formal or polite. With a friend, the tone is more informal and the use of slang is even possible. If the user is angry, the style of the conversation may be more aggressive.

The personalization of the system to a given user is initially done by storing his/her most appropriate sentences to the most predictive affective states. At the beginning, a set of sentences is stored in files or in a database, along with their possible usage (the most adequate moment to make them known, the style of the conversation, the state of mind they are related to, etc.), as mentioned in Garay-Vitoria et al. (1995). The user can include (or remove) new sentences for these states. Furthermore, sentences interactively composed during a given conversation may be stored for displaying in future conversations, whenever the need arises. An added feature is the possibility of storing proper nouns in order to personalize the user's environment, possibly with information related to the way users treat their peers (formal with the boss, informal with a brother, slang with a friend, etc.). In addition, the interlocutor's attention must be maintained while the user is interactively composing a message. In this stage, filler messages are expressed at regular intervals, minimizing awkward silences.

Moreover, Gestele has predefined profiles that may be selected in order to properly express voice characteristics by using the system (see Figure 5). There are five predefined

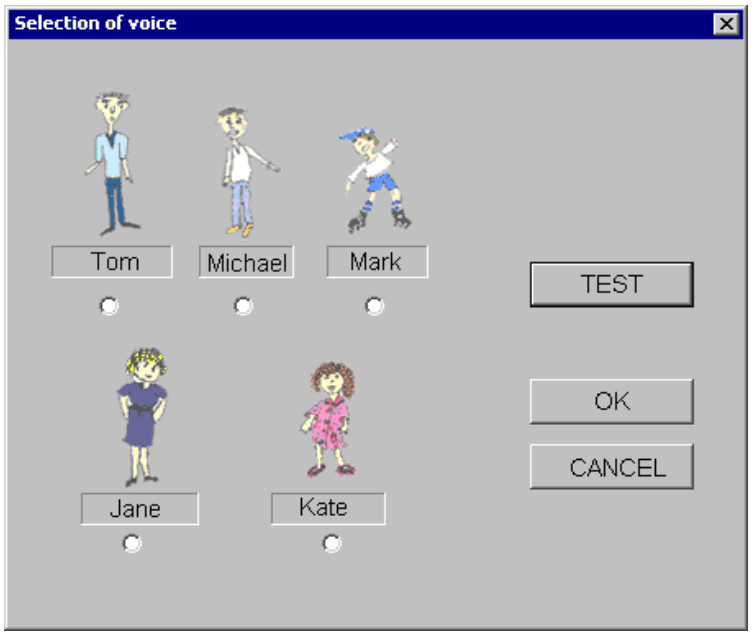

Figure 5. Selecting voice predefined characteristics in Gestele system.

profiles: male adult, male teenager, male child, female adult, and female child. Even though these profiles are predefined, the voice parameters can be adjusted using the previously mentioned voice options tool.

The interface has been designed to be usable and accessible. It is easily learnable (e.g., icons are associated with descriptive labels to avoid confusion); flexible and adaptable (e.g., selections can be made directly or via scanning input by using the keyboard, the mouse, or another input device); direct (e.g., the messages shown to users are clear and concrete); and with enough feedback (e.g., there is an acoustic signal that expresses the beginning of actions in order to make clear to the users when a new action starts). For a better understanding of the Gestele system, we provide a detailed description of a real-use scenario.

Use Scenario

Associated with the E0 state of the automaton shown in Figure 2 is the Gestele interface shown in Figure 3. At this point, according to the automaton, the user can make a call (selecting the Call button in the interface), receive a call (selecting the Answer button in the interface) and prepare a message (selecting the Compose button in the interface). Or the user also can change the configuration characteristics (relative to the voice, the scanning, or similar topics) or explicitly select his/her emotion and the conversation style. Configuration characteristics also can be changed at any moment while the application is running.

To make a call, the user selects the Call button (either directly or by scanning), and is prompted for the telephone number. A new window emerges to write the phone number (see Figure 6). In this new window, the user may directly write the number to be called or select a number from a directory in the system. The current state in the automaton is now E2, and system interface is the one shown in Figure 4.

Figure 6. Making a telephone call in Gestele system.

Next, when telephone connection is established, a salutation or notice is made, and the current state is now E4. When the main topic is introduced, the system is ready to aid in the current specific conversation (E5 state on the automaton). While the interface of the system is quite similar to Figure 4, the Greetings button is not selectable. The matrix options and the prediction list are windows to help the user compose. Windows can be closed if they are not useful (e.g., when the user directly writes messages by using a keyboard or a similar input device). After writing in the text box, sentences can be expressed via synthetic voice and sent through the telephone line by clicking the Speak button. The text box can be cleared by clicking the Delete Text button. The sentence that was just spoken is made available in the text box in order to repeat it very easily if the interlocutor does not hear it and makes a request for its repetition. Repetition is made by clicking the Speak button again. These pre-stored sentences can be selected by using the Select button at the top of the interface. After the main topic, the user will usually say a farewell (the Goodbye button). The automaton goes to E6 state and the telephone will then be hung up and the automaton returns to E0 state and the interface to Figure 3.

Once a conversation is established, the telephone can be hung up at any moment by clicking the Hang Up button. The automaton will return to E0 state and the interface will be the one shown in Figure 3.

Affectivity in Gestele

On the Gestele interface (see Figures 3 and 4), there are two buttons to reflect the user's emotion and style of conversation in which he/she is involved. We have taken four possible values for each of the topics into account. Possible user emotions are happiness, sadness, anger, and neutrality. On the other hand, possible styles of conversation are formal, informal, humorous, and aggressive. This information is used for the automatic generation of affect within the sentences. Depending on these values, the sentences and the way they are synthesized are different. Sentences are stored in a database and categorized in terms of the state of the automaton they can be used in. Sentences are also indexed by the type of emotion and conversation to select the adequate expression on a given context.

Transmission of Affective Characteristics

To reflect the user's emotion, the three parameters (volume, pitch, and speed) used by the voice synthesizers are tuned in the way proposed by Iriondo et al. (2000), Murray (2000), Murray, Arnott, & Rohwer (1996), and Oudeyer (2003) in order to emulate diverse emotions. The most appropriate type of voice for each user can be selected among a set of predetermined voices. These voices can also be adjusted to the user's preferences.

The emotions and the style of conversation change only if the user wants to do so. In this way, the user always has control over the values of these two options. Nevertheless, as reported by Picard (1997), the selection of certain affective parameters may require a tremendous effort for users with disabilities. In order to avoid interrupting the message-writing user with the task of selecting a frame of mind and/or a style of conversation, the use of automatic emotion capture methods is suitable. While the system detects the user's frame of mind, he/she can focus on message composition. Consequently, the system will automatically detect the style of conversation and emotion and adapt itself to the most

adequate characteristics of this with a minimum effort on the part of the user (who only needs to supervise the adaptation).

Recognition of the Frame of Mind and the Style of Conversation

To know about the style of conversation, both the user's emotional state and the interlocutor's frame of mind have to be considered. Concerning the automatic detection of the interlocutor's emotion, we have carried out a preliminary study on detecting the volume, pitch, and speed for interlocutors who are able speak directly. We have created a basic model based on fuzzy logic that takes the variations of those factors into account and estimates the interlocutor's emotion. We have established a knowledge base related to various interlocutors and their normal affective values. This knowledge base is used to define certain user categories and to evaluate how the changes in the interlocutor's volume, pitch, and speed are related to interlocutor's emotion. This feature will be included in a future new version of the Gestele system.

Both the user's and interlocutor's emotions have an important influence on the style of conversation. If the interlocutor is a frequent conversant, and therefore in the stored knowledge base, a particular style can be highlighted. Then, taking the user's emotions and the style of conversation once again into account, the most adequate sentences will be spoken (taken from the database of the sentences) with the parameters (volume, pitch, speed) that better express the feelings of the user with disabilities.

Even when the user's frame of mind is established, the system still gives the user the option of directly changing it (using a scrollable list). In addition, the system has a list of certain key words that may reflect the frame of mind. For example, if the user is writing insults, the system should expect the user to be angry. On the other hand, if some of the words are compliments, then the system projects that the user is happy with something or somebody. Of course, this depends on the user's personal characteristics and use of language (as not everybody uses insults and compliments in the same way and, for example, insults are sometimes used in jest). A weighted dictionary of key terms related to particular frames of mind has been built manually.

EMPIRICAL STUDY ON EMOTION TRANSMISSION VIA THE TELEPHONE

In order to apply the Gestele system in communications via telephone, we wanted to know how much the distortion introduced by the use of telephone line, both in natural and synthetic voice, would affect the emotion recognition by the interlocutor. That is, we cannot take for granted that the understanding of expressive parameters via the telephone is similar to the direct hearing of the same voice.

For this reason, we designed an experiment to assess whether the quality lost due to the use of the telephone would affect emotion recognition (Garay, Fajardo, López, & Cearreta, 2005). The TTS engine of Gestele was used to synthesize audio files with different characteristics by manipulating voice parameters. In this study, four emotional states were focused on: neutral, happy, sad, and angry. The objective was to verify whether listeners perceived differences in the understanding of these four emotions in the same phrases heard directly or over the telephone. The hypothesis was that the transmission of expressivity with telephone communication would be less efficient than that obtained in direct communication.

Method

Participants were 25 student and professor volunteers from the Computer Science faculty (University of the Basque Country, Spain), 17 males (average age 33.5 years old) and 8 females (average age 39.4 years old). This preliminary study focused on the paralinguistic parameters of the speech because the synthesized language (English) was different than the mother language (Spanish) of the volunteers. This way, the effect of the sentences' meaning was controlled. In addition, the English level of the participants was surveyed and introduced as a covariate variable in the statistical analyses. The participants' English level was classified following the Spanish standards, as elementary (12% of the sample), intermediate (56%), first certificate (24%), advanced (24%) and proficiency (4%).

Ninety-six sentences reflecting the various paralinguistic emotions were produced. A computer program to gather the results was developed.

Hardware

A Microsoft SDK 5.1 TTS engine was used to synthesize the voice in mono-aural PCM (Pulse Code Modulation). Sentences were uttered in two formats: direct voice quality was presented at 22050 Hz with 16 bits and telephone quality was simulated using 8000 Hz with 8 bits.

Design of the Experiment

A multifactor-within-subject design was adopted. The independent variables were voice type (Direct and Telephone), emotional status presented in the spoken statements (Neutral, Happy, Sad, and Angry), and a combination of three parameters values (Volume, Rate, and Pitch) for emotional status (giving as a result combinations named 1, 2, and 3 for each emotional status). The dependent variable was the rate or correspondence (in percentage) between the answer of the participants and the emotion programmed for the synthetic voice. This variable was called "hits."

A variable that could interfere in the effect of the manipulated variables is the content of the sentences. To avoid this effect, only four types of sentences were used, each reflecting neutral, happy, sad, or angry semantics (see Table 2). Additionally, each type of sentence was combined with the three possible combinations for each emotional status. In this way, this variable was neutralized and was not taken into account in the subsequent statistical analysis.

Table 2. Sentences Used in the Study.

Intention	Sentence
Happy	I enjoy cooking in the kitchen.
Neutral	Wait a moment, I am writing.
Angry	Your mother is worse than mine is!
Sad	I feel very tired and exhausted.

Procedure

Each person was asked to listen to two blocks of 48 sentences each through headphones, and to match each sentence heard with an emotional status. Sentences within each block were uttered one by one as if either directly by the synthesizer or with telephone quality. Half of the participants started the experiment with the block of direct voice sentences and the other half began with the telephone quality; the groups then listened to the alternate block. The order of presentation was randomly assigned to each participant. In the same way, to avoid any dependence, the order of presentation of each emotional status was randomly distributed within each block of sentences.

Each sentence was uttered twice, with a 1-second gap between utterances. After that, participants had to select one of the emotions (neutral, happy, sad, or angry) from a form shown on a computer screen. The next sentence wasn't voiced until the participant answered. Each sentence was presented in the same manner until all 48 sentences of each block were spoken. To ensure the comprehension of the procedure by the subjects, a trial block was carried out before the experimental phase.

Results

With the data obtained, an ANCOVA multifactorial study was performed. Thus, independent variables within subject were Type of Voice (Direct or Telephone), Emotion (Neutral, Happy, Sad, and Angry) and Combination of Voice Parameters Values (1, 2, 3; see Tables 3 and 4)[3]. The knowledge of the English language was introduced as a covaried variable. The percentage of hits was the dependent variable.

The most interesting result was that there were no significant differences in emotion perception for the voice directly heard or heard over the telephone. In addition, a significant effect of the Emotion Type variable was obtained, F (3, 72) = 18.52, MSE = 0.14, $p < 0.001$. Sad obtained M = 0.80 hits on average; Angry, M = 0.70; Neutral, M = 0.66; and Happy, M = 0.66. As seen in Figure 7, the emotions Neutral and Happy were significantly harder to detect than Sad and Angry, F (1, 24) = 416.34, MSE = 0.12, $p < 0.001$. In the same way, Sad was significantly easier to detect than Angry, F (1, 24) = 5.74, MSE = 0.13, $p < 0.001$.

Table 3. Generic Values of Synthesized Voice Characteristics.

	Volume	Rate	Pitch
Range	0/100	-10/+10	- 10/ +10
Default	DV=100	DR = 0	DP = 0
Maximum	100%	DR*3	DP*4/3
Minimum	0	DR/3	DP*3/4
Increments	1%	Rate + $\sqrt[10]{3}$	Pitch+ $\sqrt[24]{2}$
Scale	Linear	Logarithmic	Logarithmic

Table 4. Specific Combinations of Voice Parameters Used in the Study.

Emotions	Volume	Rate	Pitch	Combination
Neutral	80	0	0	Neutral 1
	85	0	0	Neutral 2
	90	0	0	Neutral 3
Happy	100	3	8	Happy 1
	80	1	10	Happy 2
	90	2	9	Happy 3
Sad	60	-4	-8	Sad 1
	45	-2	-10	Sad 2
	55	-3	-9	Sad 3
Angry	100	2	3	Angry 1
	100	3	7	Angry 2
	100	2	5	Angry 3

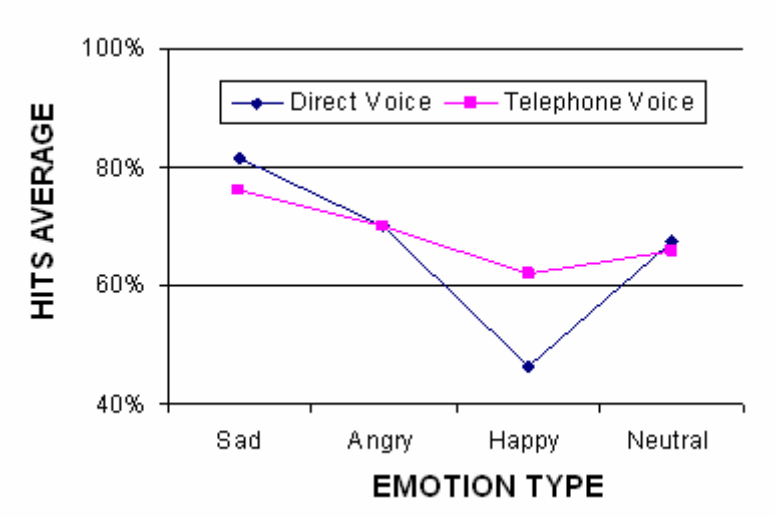

Figure 7. Hits averages (percentages of emotions recognized by users) for each type of emotion condition transmitted via direct synthetic voice or via telephone synthetic voice.

According to these results, we can conclude that the transmission of emotional cues associated with synthetic voice utterances is equally efficient whether the voice is heard over the telephone or directly. In addition, our study allowed us to partially reiterate the results obtained by Oudeyer (2003), showing the manipulation of volume, rate, and pitch parameters of synthetic voice allows for the expression of emotions. Nevertheless, there are certain emotions still difficult to reproduce, especially happiness and neutrality. The emotions of sadness and anger are perceived with better accuracy. The superiority in perceiving an angry expression seems to agree with the results obtained with human voices (Johnstone & Scherer, 2000). In that study, the authors suggest an evolutive explanation: Emotions that express danger, such as anger and fear, must be able to be communicated large distances with the aim

of being perceived accurately by the members of the group or by the enemies. In order to do so, voice is the most effective means (as the results reveal), while facial gesture would be more effective for emotions that must be transmitted short distances.

These results must be considered with caution, as the experiment did have several methodological limitations. One of the most important was the lack of comparison with the efficiency of human voice, as both direct and telephone voices were synthetic.

CONCLUSIONS AND FUTURE WORK

Even considering all the advances reached in this area, for computers to recognize, identify, and synthesize emotions in real life remains science fiction at the moment. Nevertheless, this is a possibility that is being refined today. As shown, the application of affective techniques in assistive technology systems to enhance the rehabilitation of, integration of, and communication among people with disabilities is very promising. Most of the efforts so far have been centered in unimodal and unistage interaction; however, there are also studies related to multimodal interactions.

In particular, the work presented in this paper shows that affective mediation may serve to transmit information through mediation systems. This therefore enhances the user's expressive features (by text and speech) and his/her ability to recognize and model messages.

Gestele adds information related to the user's affects and to the style of conversation even over the telephone, thanks to the associated speech synthesizer that modifies various basic prosodic parameters. In order to minimize the effort required by the users to inform their interlocutors about their current emotional state, the model has been designed to adapt automatically, depending on the words used in the conversation. Such a modeling component must be improved and tested with end users (mobility- and speech-impaired people) in the future. Furthermore, other ways for the automatic recognition of user affect are being designed within the many projects LHCISN is involved in. The above-mentioned alternative ways will embrace the rest of the systems of the Lang's (1984) bio-information model, that is, behavioral (facial gestures, speech prosody, etc.) and physiological responses (heart rate, skin conductance, etc.).

Finally, we are working on developing a decision model that allows for integrating and interpreting information taken from diverse sources (e.g., speech, facial gesture, and heart rate), as found in Obrenovic, Garay, López, Fajardo, & Cearreta (2005). Such a significant step is a huge challenge, as little work has been done on how the three response systems described by Lang (verbal, behavioral, and physiological) interact among themselves. From another point of view, the applied research itself may make the basic research easier.

ENDNOTES

1. See, for example, Tegic Communications, T9 ® text input for keypad devices. Retrieved September 26, 2005, from http://www.tegic.com
2. See, for example, BodyMedia, Bio-metric monitoring system. Retrieved September 26, 2005 from http://www.bodymedia.com/products/biotransceiver.jsp

3. This variable was introduced as the recommended ranges of each parameter, for each emotional state is wide and variations in the selected values or combination of values could affect the efficiency of the comprehension of such an emotional state.

REFERENCES

Affecting Computing. (n.d.). Retrieved September 26, 2005, from http://affect.media.mit.edu/

Alegría, J. (1999). La lectura en el niño sordo: Elementos para una discusión [Reading in the deaf child: Elements for discussion]. In A. B. Domínguez & C. Velasco (Eds.), *Lenguaje escrito y sordera: Enfoques teóricos y derivaciones practices* (pp. 59–76). Salamanque, Spain: Pontifical University of Salamanca Press.

Alm, N., Arnott, J. L., & Newell, A. F. (1992). Prediction and conversational momentum in an augmentative communication system. *Communications of the ACM*, *35*(5), 46–57.

Arruabarrena, A., & González-Abascal, J. (1988). Comunicadores y emuladores de teclado [Keyboard communicators and emulators]. In M. Fernández de Villalta (Ed.), *Tecnologías de la información y discapacidad* (pp. 133–156). Madrid, Spain: FUNDESCO Ed.

Arnold, P., & Mills, M. (2001). Memory for faces, shoes and objects by deaf and hearing signers and hearing nonsigners. *Journal of Psycholinguistic Research*, *30,* 185–195.

Arnott, J. L., & Javed, M. Y. (1992). Probabilistic character disambiguation for reduced keyboards using small text samples. *Alternative and Augmentative Communication, 8,* 215–223.

Asensio, M. (1989). *Los procesos de lectura en los deficientes auditivos* [Reading processes in the auditory deficients]. Unpublished doctoral dissertation. Autonomous University of Madrid, Spain.

Birkby, A. (2004). *Emotional Hearing Aid: An assistive tool for individuals with Asperger's syndrome*. Diploma project dissertation, St. Catharine's College. Retrieved September 26, 2005, from http://www.cl.cam.ac.uk/~re227/completed_projects/alex-dissertation.pdf

Boucher, J., & Lewis, V. (1992). Unfamiliar face recognition in relatively able autistic children. *Journal of Child Psychology and Psychiatry*, *33*, 843–859.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW). Stimuli, instruction manual and affective ratings* (Tech. Rep. C-1) Gainesville: University of Florida, The Center for Research in Psychophysiology.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*, 119–155.

Capps, L., Yirmiya, N., & Sigman, M. D. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Journal of Child Psychology and Psychiatry*, *33*, 1169–1182.

Casacuberta, D. (2001). *La mente humana: Diez enigmas y 100 preguntas* [The human mind: Ten enigmas and 100 questions]. Barcelona, Spain: Océano Ed.

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, *29*, 57–66.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, *18*(1), 32–80.

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & De Carolis, B. (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies, 59,* 81–118.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology*, *115*, 107–117.

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologist Press.

Fajardo, I. (2005). *Cognitive accessibility to hypertext systems: The role of verbal and visuospatial abilities of deaf and hearing users in information retrieval*. Unpublished doctoral dissertation, University of Granada, Spain.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is special about face perception? *Psychological Review*, *105*, 482–498.

Foulds, R. A., Soede, M., & Van Balkom, H. (1987). Statistical disambiguation of multi-character keys applied to reduce motor requirements for Augmentative and Alternative Communication. *Alternative and Augmentative Communication*, *3*, 192–195.

Foundation for Assistive Technology, (n.d.). Retrieved February 15, 2006, from http://www.fastuk.org/

Garay, N., & Abascal, J. (1994). Using statistical and syntactic information in word prediction for input speed enhancement. In C. Chrisment (Ed.), *Information Systems Design and Hypermedia* (pp. 223–230). Toulouse, France: Cépaduès-Éditions.

Garay, N., Abascal, J., & Gardeazabal, L. (2002). Mediación emocional en sistemas de Comunicación Aumentativa y Alternativa [Emotional mediation in Augmentative and Alternative Communication systems]. *Revista Iberoamericana de Inteligencia Artificial*, *16*, 65–70.

Garay, N., Fajardo, I., López, J. M., & Cearreta, I. (2005). Estudio de la transmisión de emociones mediante voz sintética por vía telefónica [Empirical study in emotion transmission by means of synthetic speech by via telephone]. In Á. R. Puerta & M. Gea (Eds.), *Proceedings of the 6th Congress of Human Computer Interaction* (Interacción 2005; pp. 3–10). Granada, Spain: Thomson Ed.

Garay-Vitoria, N. (2001). *Sistemas de predicción lingüística: aplicación a idiomas con alto y bajo grado de flexión, en el ámbito de la Comunicación Alternativa y Aumentativa* [Linguistic prediction systems: Application to languages with high and low inflection levels, in the scope of Alternative and Augmentative Communication]. Leioa, Spain: University of the Basque Country Press.

Garay-Vitoria, N., Abascal, J., & Gardeazabal, L. (2001). Adaptive emotional interface for a telephone communication aid. In C. Marincek, C. Bühler, H. Knops, & R. Andrich (Eds.), *Assistive Technology: Added Value to the Quality of Life* (pp. 175–179). Ljubljana, Slovenia: IOS Press & Ohmsa.

Garay-Vitoria, N., Abascal, J. G., & Urigoitia-Bengoa, S. (1995, March). *Application of the human conversation modelling in a telephonic aid*. Paper presented at the 15th International Symposium on Human Factors in Telecommunications (HFT '95), Melbourne, Australia.

Gardeazabal, L. (2000). *Aplicaciones de la tecnología de computadores a la mejora de la velocidad de comunicación en sistemas de Comunicación Aumentativa y Alternativa* [Applications of computer technology to the improvement of the speed communication in Augmentative and Alternative Communication systems]. Leioa, Spain: University of the Basque Country Press.

Gershenfeld, N. (2000). *When things start to think*. New York: Henry Holt & Company Press.

Hall, E. T. (1998). The Power of Hidden Differences. In M. J. Bennett (Ed.), *Basic Concepts of Intercultural Communication: Selected Readings* (pp. 53–67). Yarmouth, ME: Intercultural Press, Inc.

Hall, G. B. C., Szechtman, H., & Nahmias, C. (2003). Enhanced salience and emotion recognition in autism: A pet study. *American Journal of Psychiatry*, *160*, 1439–1441.

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction. *International Journal of Human–Computer Studies*, *59*, 1–32.

International Society for Augmentative and Alternative Communication [ISAAC]. (n.d.). Retrieved February 07, 2006, from http://www.isaac-online.org/en/aac/what_is.html

Iriondo, I., Guaus, R., Rodríguez, A., Lázaro, P., Montoya, N., Blanco, J. M., Bernadas, D., Oliver, J. M., Tena, D., & Longhi, L. (2000, September). *Validation of an acoustical modelling of emotional expression in Spanish using speech synthesis techniques*. Paper delivered at the ISCA [International Speech Communication Association] Workshop on Speech and Emotion, Belfast, Northern Ireland. Retrieved September 26, 2005, from http://www.qub.ac.uk/en/isca/proceedings/pdfs/iriondo.pdf

Jacko, J. A., Vitense, H. S., & Scott, I. U. (2003). Perceptual impairments and computing technologies. In J. A. Jacko & A. Sears (Eds.), *The human–computer interaction handbook* (pp. 504–522). Mahwah, NJ: Erlbaum.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.). *Handbook of emotion* (2nd ed.; pp. 220–235). New York: Guilford Publications.

Kaliouby, R., & Robinson, P. (2005). The emotional hearing aid: An assistive tool for children with Asperger's syndrome. *Universal Access in the Information Society, 4,* 121–134.

Klin, A., Sparrow, S., De Bildt, A., Cicchetti, D., Cohen, D., & Volkmar, F. (1999). A normed study of face recognition in autism and related disorders. *Journal of Autism and Developmental Disorders*, *29*, 499–508.

Knapp, M. L. (1980). *Essentials of nonverbal communication*. New York: Holt, Rhinehart & Winston.

Lang, P. J. (1979). A bio-informational theory of emotional imagery. *Psychophysiology, 16,* 495–512.

Lang, P. J. (1984). Cognition in emotion: Concept and action. In C. Izard, J. Kagan, and R. Zajonc (Eds.), *Emotions, cognition and behavior* (pp. 192–226). New York: Cambridge University Press.

Lang, P. J. (1995). The network model of emotion: Motivational connections. In R. S. Wyer & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 6; pp. 331–352). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesher, G. W., Moulton, B. J., & Higginbotham D. J. (1998a). Optimal character arrangements for ambiguous keyboards. *IEEE Transactions on Rehabilitation Engineering*, *6*, 415–423.

Lesher, G. W., Moulton, B. J., & Higginbotham, D. J. (1998b). Techniques for augmenting scanning communication. *Augmentative and Alternative Communication*, *14*, 81–101.

Levine, S. H., Goodenough-Trepagnier, C., Getschow, C. O., &. Minneman S. L. (1987). Multi-character key text entry using computer disambiguation. In R. D. Steele, W. Gerrey (Eds.), *Proceedings of the 10th Annual Conference on Rehabilitation Engineering* (pp. 177–179). San Jose, CA: RESNA Press.

Leybaert, J., Alegria, J., & Morais, J. (1982). On automatic reading processes in the deaf. *Cahiers de Psychologie Cognitive, 2*, 185–192.

Lisetti, C. L., Nasoz, F., Lerouge, C., Ozyer, O., & Alvarez, K. (2003). Developing multimodal intelligent affective interfaces for tele-home health care. *International Journal of Human–Computer Studies, 59,* 245–255.

Magnuson, T. (1995). Word prediction as linguistic support for individuals with reading and writing difficulties. In I. Placencia Porrero & R. Puig de la Bellacasa (Eds.), *The European context for assistive technology: Proceedings from the 2nd TIDE Congress* (pp. 320–323). Amsterdam: IOS Press/Ohmsa.

Mehrabian, A. (1971). *Silent Messages*. Belmont, CA: Wadsworth Publishing Co.

Moshkina, L. & Moore, M. (2001, July). *Towards affective interfaces for neural signal users*. Workshop on Attitudes, Personality and Emotions in User-Adapted Interaction, in conjunction with the User Adaptation and Modeling Conference, Sonthofen, Germany**.**

Murray, I. R. (2000). Emotion in concatenated speech. In *Proceedings of the IEE Seminar State of the Art in Speech Synthesis* (pp. 7/1–7/6). London: IEE Press.

Murray, R., Arnott, J. L., & Rohwer, E. A. (1996). Emotional stress in synthetic speech: Progress and future directions. *Speech Communication, 20,* 85–91.

Noyes, J. M., & Frankish, C. R. (1992). Speech recognition technology for individuals with disabilities. *Augmentative & Alternative Communication*, *8*, 297–303.

Obrenovic, Z., Garay, N., López, J. M., Fajardo, I., & Cearreta, I. (2005). An ontology for description of emotional cues. In J. Tao, T. Tan, & R. W. Picard (Eds.), *Proceedings of the First International Conference on Affective Computing & Intelligent Interaction* (ACII'05; pp. 505–512). Beijing, China: Springer.

Oregon Health & Science University [OHSU]. (2005). Researchers study communication disorders in autism. Retrieved February 07, 2006, from http://www.ohsu.edu/ohsuedu/newspub/releases/022805autismhtml.cfm

Oudeyer, P.-Y. (2003). The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies, 59*, 157–183.

Pelpherey, K., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*, 249–261.

Petrushin, V. A. (2000). Emotion recognition agents in real world. In K. Dautenhahn (Ed.), *Socially intelligent agents: The human in the loop* (pp. 136–138). Menlo Park, CA: AAAI Press.

Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Robinson, G. L., & Whiting, P. J. (2003). The interpretation of emotion from facial expression for children with visual processing problems. *Australasian Journal of Special Education*, *27*(2), 50–67.

Roy, S. (2003). State of the art of Virtual Reality Therapy (VRT) in phobic disorders. *Psychology Journal*, *1*, 176–183.

Schröder, M. (2003). *Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis*. Unpublished doctoral thesis, Institute of Phonetics, Saarland University, Germany.

Silvestre, N., & Valero, J. (1995). Investigación e intervención educativa en el alumno sordo: cuestiones sobre la integración escolar [Research and educative intervention in the deaf student: Questions about school integration], *Infancia y Aprendizaje*, *69-70*, 31–44.

Strapparava, C., & Valitutti, A. (2004, May). *WordNet-Affect: An affective extension of WordNet*. Paper presented at the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal.

Tantam, D., Monaghan, L., Nicholson, H., & Stirling, J. (1989). Autistic children's ability to interpret faces: A research note. *Journal of Child Psychology and Psychiatry*, *30*, 623–630.

Tao, J., & Tan, T. (2005). Affective computing: A review. In J. Tao, T. Tan, & R. W. Picard (Eds.), *Proceedings of the First International Conference on Affective Computing & Intelligent Interaction* (ACII'05; pp. 981–995). Beijing, China: Springer.

Vanderheiden, G. C. (1998). Universal design and assistive technology in communication and information technologies: Alternatives or complements? *Assistive Technology*, *10*, 29–36.

VanLancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotion-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, *5*, 207–226.

Venkatagiri, H. S. (1999). Efficient keyboard layouts for sequential access in Augmentative and Alternative Communication. *Augmentative and Alternative Communication*, *15*, 126–134.

Zordell, J. (1990). The use of word prediction and spelling correction software with mildly handicapped students. *Closing the Gap*, *9*, 10–11.

Authors' Note

The involved work has received financial support from the Department of Economy of the local government "Gipuzkoako Foru Aldundia."

All Correspondence should be addressed to:
Nestor Garay
Laboratory of Human–Computer Interaction for Special Needs
Computer Science Faculty, University of the Basque Country
Manuel Lardizabal 1; E-20018 Donostia (Gipuzkoa). Spain
e-mail: nestor.garay@ehu.es