NAIST-IS-DD0361216

Doctoral Dissertation

Study on the Impressions of Robot Bodily Expressions on the Human Observer

Khiat, Abdelaziz

March 23, 2007

Department of Information Systems
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Khiat, Abdelaziz

Thesis Committee:

Professor Tsukasa Ogasawara	(Supervisor)
Professor Ken-ichi Matsumoto	(Co-supervisor)
Associate Professor Yoshio Matsumoto	

To my family

Study on the Impressions of Robot Bodily Expressions on the Human Observer*

Khiat, Abdelaziz

Abstract

Recently, robotics research has investigated issues surrounding the interaction modalities with robots, how these robots should look and how their behavior should adjust while interacting with humans. It is believed that in the near future robots will be more prevalent in human environments. Thus it is important to understand accurately our reactions and dispositions toward robots in different circumstances. Moreover, the robot's correct production and perception of social cues is very important. Humans have developed advanced skills in interpreting the intentions and bodily expressions of other human beings. If similar skills were possessed by robots, it would allow them to generate behaviors that are familiar to us and thus increase their chances of being accepted as partners in our daily lives.

This dissertation deals with the assessment, using brain activity, of the kind of impression an individual has when observing a particular type of robot bodily expressions. Using brain signals or electroencephalogram (EEG) to control devices is one form of Brain-Machine Interfacing (BMI). A BMI system is usually targeted for applications intended for handicapped and/or aged people. In our case, we used brain activity for evaluation purposes to characterize specific properties of bodily expressions that have a noticeable impact on an observer. First, a number of bodily expressions were generated and executed by a humanoid

*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0361216, March 23, 2007.

robot. Using quantitative descriptions of Laban features of Shape and Effort, the generated bodily expressions were classified into two categories. The classification was confirmed statistically with the results of a self-reporting experiment. In another experiment, the impact on brain activity of each category of bodily expressions was confirmed using spectral analysis by showing that the power level of low-alpha (8–11Hz) frequency band changes according to the category. The most reactive electrode positions were found to be those covering activity above the superior temporal sulcus (STS) and the prefrontal cortex (PFC). This supports previous research findings about the activation of mirror neurons and the STS during the perception of biological motions, learning and imitation. The repeatability of this reaction was proven with the results of a third experiment, where a subject was asked to observe the same bodily expression several times. Finally, we proposed a method that uses self-organizing maps to represent and generalize carefully selected features of brain activity in order to assess the impression of robot bodily expressions on an observer. The recognition rate achieved with this method was close to 80% when using data from a single subject. However, this rate decreased significantly when data from several subjects were used, suggesting the existence of differences in brain activity between individuals. To overcome this problem an improvement of the method was proposed, allowing an almost constant recognition rate of about 85% regardless of the number of data sources considered. To complete this study, brain activity measured during the observation of bodily expressions of a human was used to assess the observer's impression using the same method.

Keywords:

Robot Bodily Expressions, Impressions, Human-Robot Interaction, Electroencephalogram, Laban Features

ロボット身体動作表現が人に及ぼす影響に関する研究*

キアット　アブデラズィズ

内容梗概

近年，ロボット研究の分野においては，人とロボットがインタラクションする際に，ロボットが人に対してどのような印象を与えるべきか，またロボットの所作が，インタラクション中にどのように変化していくべきかなど，インタラクションモダリティに関する研究が盛んに行われている．近い将来，ロボットが我々の生活環境に数多く存在することになると予想される．そのため様々な状況において人がロボットに対してどのように反応し，どのように感じるのかを正確に理解することが重要になる．さらに，社会的な刺激 (Social Cues) をロボットが的確に提示することや，認識することも，非常に重要な課題である．人がもつ他者の意図や表現を解釈する能力は，非常に高い．ロボットとのインタラクション時，このような高機能解釈の実現が可能になれば，ロボットに自然で親しみやすい振る舞いを生成出来るようになり，そしてそれはロボットが我々の日常生活においてパートナーとして活躍できるという可能性を増大させる．

本論文は，ロボットの身体動作表現により，それぞれの人が受ける印象を脳活動から評価することを目的としている．すなわち，人に対して顕著なインパクトを与える身体動作表現が，どのような特徴量を持っているかを評価するために，脳活動を用いることを提案する．最初に，ヒューマノイドロボットを用いて，いくつかの身体動作表現の生成と実装を行った．生成されたロボットの身体動作表現を，身体動作の全身的な幾何学的特徴 (Shape) と，時間的特徴・空間的特徴・エネルギ的特徴の 3 つの要素 (Effort) とを定量的に表現した “ラバンの身体運動特徴量” を用いて，２つのカテゴリに大別した．この分類結果は，複数の被験者による印象評価実験のアンケート結果を，統計的に処理することによって，正し

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 博士論文, NAIST-IS-DD0361216, 2007年 03月 23日.

い分類であることが確認できた．さらに，それぞれの身体動作表現を観測した際の，被験者の脳波のスペクトル解析を行った．その結果，各カテゴリの身体動作表現が脳活動に与える影響の違いは，α波の低い周波数帯域のパワーの変化の違いに現れることが確認出来た．このような感覚運動皮質上での脳活動を計測対象とする際，最も良く反応し計測に適した電極位置を特定することも出来た．これは，認知プロセスにおける上側頭溝上の活動に関する従来研究の結果を，サポートするものでもある．最後に，観測者がロボットの身体動作表現から受ける印象を判断するために選ばれた脳活動の特徴をマッピングし，一般化するために，自己組織化マップ (SOM) を用いる手法を提案した．SOM を一名の被験者データを用いて作成した場合，認識率は 80% 近くに達した．しかし，複数の被験者のデータを用いた場合，この認識率は著しく低下することもわかった．そこで，階層化した SOM である GHSOM を用いた．これにより脳活動の個人差に対応でき，一定の認識率を得ることができた．補足実験として，人が演者となり身体動作表現を行った際の被験者の脳活動を計測分析した結果，ロボットで行った実験と同様の結果が得られた．

キーワード

ロボットの身体動作表現，印象，ヒューマン・ロボット・インタラクション，脳波，ラバン特徴

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Motivations

Human face-to-face interaction is complex and involves understanding the interchanges and synchronizations across multiple communication channels and time scales. Speech, gaze, facial expressions and gestures are all taking part in the interaction in a parallel and interdependent way. The form of each of these modalities is very crucial in conveying a meaning despite the differences in their manifested information. Also of great importance is the tight synchrony between modalities and the choice of the right one to use, at a certain time, for certain duration. It is known that the semantic of a message is distributed over the modalities such that the communicative intentions of the person are interpreted by combining them together.

Porting these abilities into machines has turned out to be a very difficult task. A great effort is made in this direction to provide computers and robots with the ability to engage a human in an interesting, relevant conversation with appropriate speech and meaningful gestures. The need for this ability to be integrated in machines is justified by the prediction of the prevalence of machines in our daily lives in the future. The properties or abilities that should be considered include but are not limited to:

- embodiment, whether virtual or physical.
- recognition of verbal and non-verbal input.
- generation of verbal and non-verbal output.
- use of turn-taking and feedback mechanisms.

The embodiment of intelligence has always been a superior goal in artificial intelligence research, since it is more natural and intuitive to interact with agents that can be seen and touched. From a communicative point of view, as important as the ability to recognize speech may seem, the ability to recognize non-verbal behavior is equally so. Being able to interpret the human behavior will open channels of useful information. On the other hand, the feedback and the way to present information to the user should include the ability to synthesize speech and to gesture like humans while making purposeful facial expressions. Finally, being able to understand the effect of each feedback modality on the interacting human partner is of particular interest. Because, it is often difficult to interpret and infer intentions of living creatures who might have different ways of conveying and interpreting social signals. Most such misinterpretations have no consequences, but in the interaction they could be disastrous. Humans and many other animals depend heavily on the correct production and perception of facial and bodily expressions to signal threat, submission and other information. This type of interaction is unidirectional and commonly known as monad interaction.

Each of the mentioned requirements represents a challenging problem and much has to be done before realizing communicative agents that rise to the expected level of performance. There is a great deal of research work on many of them; however the perception by humans of non-verbal information and particularly bodily expressions did not have the attention that it deserves in the context of human-robot interaction.

Figure 1.1 shows the internal parts involved during a monad interaction between a human and a robot in both directions. The monad interaction going from the robot to the human is particularly interesting. Although it represents

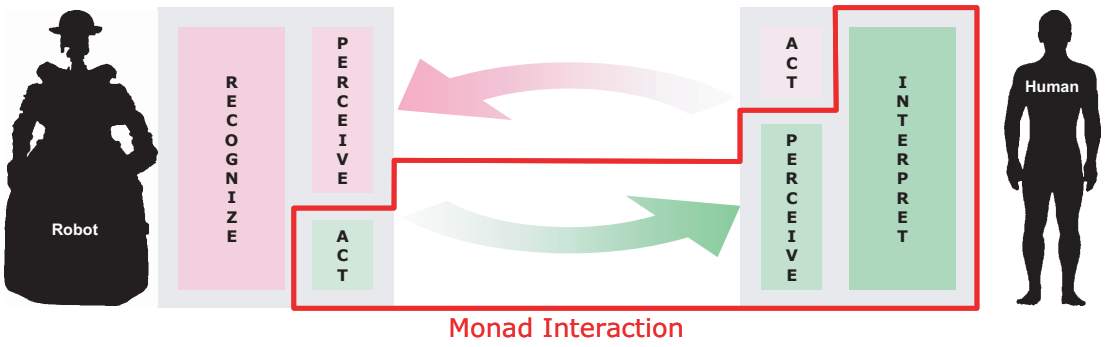


Figure 1.1. The internal parts involved during a monad interaction between a human and a robot in both directions. The highlighted monad interaction from a robot to a human is thoroughly studied in this thesis.

half of a normal bidirectional interaction (dyad interaction) it has not been considered and studied in the past. The monad interaction in the other direction, from the human to the robot is the one that has attracted the attention of most of the researchers in human-robot interaction. In order to study the considered monad interaction, there is a need to estimate the perception and interpretation processes taking place on the human side because there is no way to observe them directly. Moreover, these two processes are thought to be taking place within the human's brain, so it becomes necessary to interpret brain activity in this investigation.

1.2. Research goal

In this dissertation, I propose to explore the last issue about the effect of feedback modalities on humans. More precisely, I investigate the impression a human gets when observing certain types of robot bodily expressions and I identify the relation that exists between bodily expressions and their impressions. Since impression is by nature a subjective factor, its evaluation or assessment needs to rely on objective empirical data. I chose to reveal credible features from the observer's brain activity. This choice is motivated by the fact that all cognitive

processes take part within the human brain, thus monitoring brain activity is a reasonable decision for this investigation.

This work deals with the assessment, using brain activity, of the kind of impression an individual gets when observing a certain type of robot bodily expressions. I use brain activity for evaluation purposes to characterize specific properties of bodily expressions that have a noticeable impact on the observer. This study has clear implications for the neural basis of robot action perception and for a better understanding of critical features necessary to improve the successful development of interactive robots.

1.3. Dissertation layout

This chapter presents the motivations of this research work, along with an overview of the approach and the main goal. It explains where this work is situated within the field of human-machine interaction and its necessity.

The second chapter presents robot bodily expressions, their modeling, their generation and the quantification of their expressiveness. An experimental study that explains the relation between bodily expressions and their impressions on the observer is described and important features that reflect such impressions on the observer's brain activity are identified.

The third chapter introduces the principal technique adopted for the assessment, from brain signals, of the impressions of bodily expressions on the observer, regardless of whether they were performed by robots or humans. An improvement of this technique is presented and its performance is evaluated thoroughly. This opens the door for reflections about the types of bodily expressions and how they should be chosen to improve the interaction between robots and humans.

Finally, a conclusion terminates this dissertation by summarizing the main contributions and giving directions for future developments.

Chapter 2

Robot Bodily Expressions

2.1. Introduction

In recent years, robotics research has investigated issues surrounding the interaction modalities with robots, how these robots should look and how their behavior should adjust while interacting with humans. It is believed that in the near future robots will be more prevalent in human environments. Thus it is necessary to understand accurately our reactions and dispositions toward robots in different circumstances [43]. Moreover, the robot's correct production and perception of social cues is very important. Humans have developed advanced skills in interpreting the intentions and the bodily expressions of other human beings. If similar interpretations were possible while interacting with robots, it would allow these robots to generate behaviors that are familiar to us and thus increase their chances of being accepted as partners in our daily lives.

In this chapter, I report a study on the relation between expressions and their impacts on the observer. I also attempt to understand the effect that expressions have on the observer's brain activity. Its sensitivity to bodily expressions can be used during an interaction task since it is the center of every cognitive and emotional effort. I conducted an experimental study where several users were asked to observe different robot bodily expressions while their brain activity was recorded. The results suggest the existence of an association between the type

of bodily expressions and the change in the power level of a particular frequency band of brain activity.

2.2. Related research

The approach adopted in this study is related to three apparently distinct fields but are very much interdependent:

- human-robot interaction,
- decoding of brain activity (brain-machine interfaces),
- and social perception from the cognitive science perspective.

Each one of these fields has developed independently from the other ones. However, in recent years it became apparent that there is a need to combine these scattered individual efforts to take advantage of their complementarity.

In the following, I will introduce a short overview and the recent achievements of each of these fields in the context of this study. After that, I will explain how this study relates to these fields, and how it takes advantage of their respective accomplishments.

2.2.1 Human-robot interaction

The expressiveness of a gesture is of great significance during an interaction process. We are often required to give special attention to these signs in order to keep track of an interaction. Humans have learned to adapt their behavior and to react to positive and negative bodily expressions [5]. Although there has been remarkable work on the design issues of sociable robots [7] and affective autonomous machines [44], there has not been much work on investigating the real impact of robots bodily expressions on the human user in the context of robot-human interaction. Moreover, computer-animated characters have been used to evaluate human perception of the significance of gestures. However, animated

characters and embodied ones should be treated differently since the latter are tangible entities [54]. Knowing the effect of a generated gesture, a robot can select more accurately the most appropriate action to take in a given situation. The most direct approach to study the effect on an observer that watches a bodily expression is to use reliable physiological measurements. The brain happens to be the organ most involved in the interpretation of observed events. Thus, analyzing measured brain activity is necessary for this study.

A new research field that tries to interpret brain activity in relation to different external stimuli is called Brain-Machine Interface, an overview is presented in the next section.

2.2.2 Decoding of brain activity

Brain-machine interfaces detect brain activity and translate it into forms that are human understandable and thus could be used to produce actions of some kind, such as moving a robotics arm, moving a cursor on a screen or controlling a TV set. This technology is motivated by its potential application for handicapped people who cannot use their arms normally. However, it can also be used in reverse direction such that it inputs a specific signal in the human's brain, to enhance perceptual or motor capabilities. This is widely used with patients suffering movement disorders (Parkinson disease) to allow the motor cortex in their brains to regain their normal metabolic rhythms. It has also been successfully used to allow blind people to regain a certain capability to see the outside world by connecting the simplified output of a camera directly to the visual cortex. Another kind of BMIs is the one that only tracks brain activity without giving any feedback. This type of BMIs is used in medical diagnosis to monitor the functioning of patient brains and is also used in research investigations to study the cognitive properties of humans. There are two categories of techniques which are used in BMIs: invasive techniques that implant electrodes directly in the brain and non-invasive techniques that use sensors to scan the brain activity. **Figure 2.1** presents the mostly used techniques plotted on a two dimensional

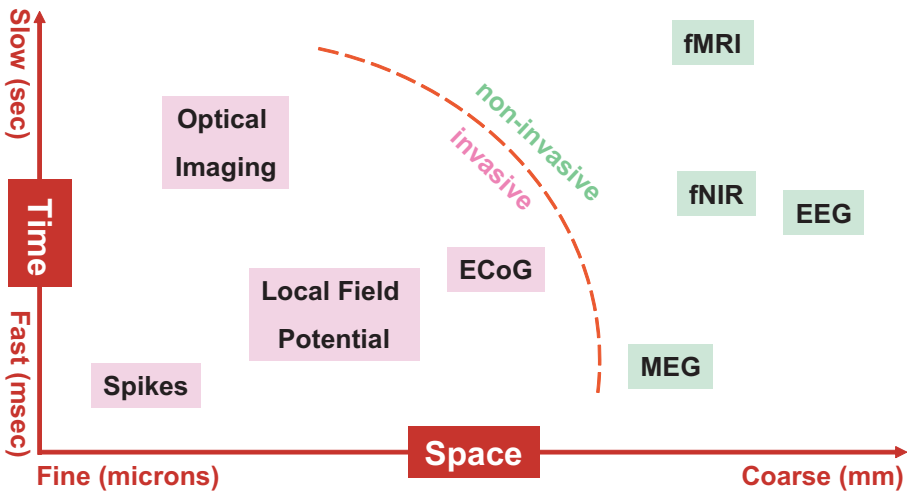


Figure 2.1. Different techniques to measure brain activity (ECoG: ElectroCortecography, fMRI: functional magnetic resonance imaging, fNIR: functional near infrared imaging, EEG: Electroencephalography, MEG: Magnetoencephalography).

space according to their spatial and temporal resolutions. It can be noticed that invasive techniques have better spatial resolution. However, they require surgery and carry the risk of infection or brain damage. On the other hand non-invasive techniques are less intrusive. However, they have a coarse spatial resolution and thus measure brain activity less precisely.

In the past, a lot of research work investigating brain activity was done on monkeys. Nicoleilis team at Duke University [63] developed a system that allowed a monkey to control a robotic arm to achieve the task of reaching a tray and carry food to its mouth. They used the very invasive technique of local field potentials to measure the activity of the motor cortex and use it to estimate the 3D trajectory of the robotic arm. In order to replace the electrodes, surgery was necessary every one or two weeks, making this technique only useful for research investigations. Wolpaw team at New York state university [66] developed a method that allowed users to control movements of a cursor on a screen. They used non-invasive EEG caps to acquire the signals of brain activity.

The drawback of this method is that the users had to go through long training sessions to achieve an acceptable level of control and learning speed differed from one user to another. Leuthhardth at the same university [38] used invasive ECoG measurements to record signal patterns corresponding to simple tasks such as hand opening and word pronouncing. These actions are recognized later on by an online comparison of the measured brain activity with the previously recorded patterns. In this work, they could also get the same result when the patient only imagined doing one of the actions previously cited, suggesting that it is possible to use this method in prosthesis for handicapped people. Probably the most successful BMI so far is the BrainGate developed by Donoghue team [20] at Brown university. They implanted a tiny micro-electrode array in the motor cortex of a handicapped patient. This one was able to control a TV set, to move a cursor on a screen, to write emails and to control a robotic arm. However, this approach requires surgery and the replacement of the electrode array regularly after that. It also needs a training for several days before achieving the required performance. Another non-invasive BMI uses NIRS measurements of train driver's brain activity [56]. This method allowed the estimation of whether the driver was using the manual or the automatic driving modes. The goal is to estimate when the driver is more prone to error, since train driving is more monotonous than car driving, thus the human error risk is higher.

Several BMIs have been designed for different reasons. In my case, I want to study the effect of bodily expressions on the brain, so I am more interested in brain activity reaction to these particular expressions and social cues in general.

2.2.3 Social perception and brain activity

Neuroimaging and neurophysiological studies in humans, suggest that early stages in the analysis of bodily movement are instantiated in specific brain regions near the superior temporal sulcus of both hemispheres [1] and near the sensory-motor cortex [45]. **Figure 2.2** illustrates the perception path during observation of biological motions. From the visual cortex the posture and the motion are treated

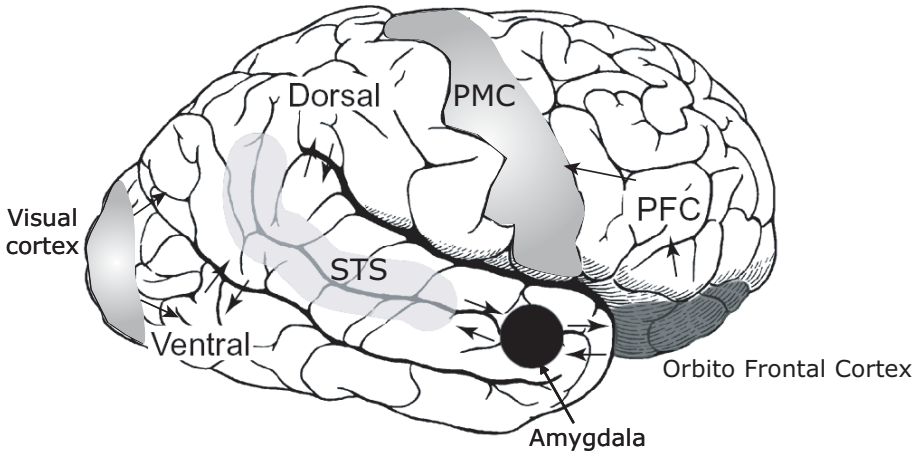


Figure 2.2. Perception path during the observation of biological motions (STS: Superior temporal sulcus, PFC: Prefrontal cortex, PMC: Primary motor cortex) *(Adapted from [1])*

separately on the dorsal and ventral parts of the brain. After that, the information goes to the superior temporal sulcus (STS), which sends this information to the orbito-frontal cortex (OFC) through the amygdala. The OFC then sends the information to the prefrontal cortex (PFC), which is connected to the primary motor cortex (PMC) and the basal ganglia, thus closing the loop from perception to action.

Within this path, some specific parts are activated when observing social cues and during imitation and learning tasks. More specifically, there were reports about the reaction of specific brain regions to eye gaze [65, 48, 21], hand action [15, 6], hand movement [16], hand grasp [50, 13], mouth movement [48, 47], and finally body movement [6, 23, 36, 17].

It is then conceivable to study the effect of social cues and particularly bodily expressions on the brain activity of humans. Cognitive science tells us which parts of the human brain to observe. Brain activity decoding techniques would help us to find the proper methods to use for brain signal interpretation, the result of which would open new application opportunities in the field of human-robot interaction.

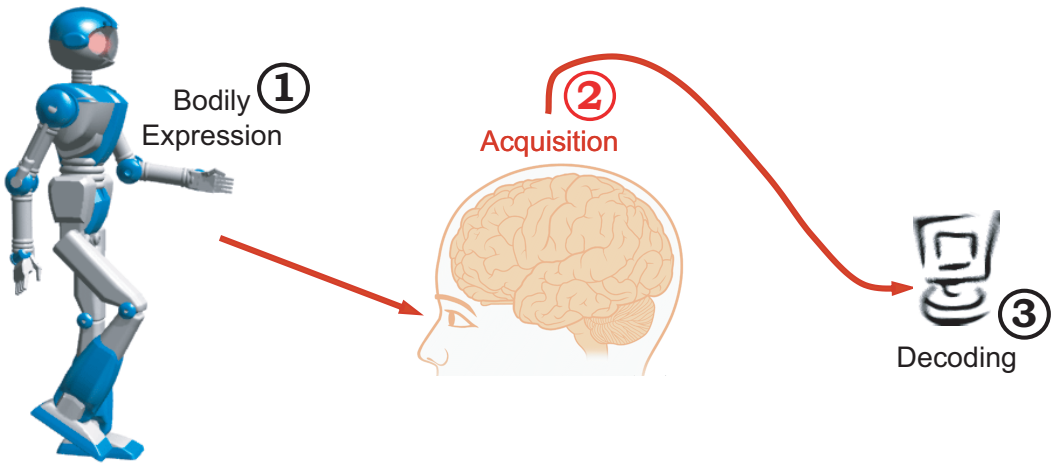


Figure 2.3. Considered scenario for robot bodily expression and its perceived impression.

2.3. Bodily expressions and their impressions

The considered scenario for this study is depicted in **Figure 2.3**. First, we have a robot that is executing a series of movements. It transmits to the observer a meaningful expression which is called *bodily expression* ①. Second, we have a human observer that perceives the expression and interprets it using his/her *a priori* knowledge ②. Then, the observer gets an *impression*, which means that bodily expression affects him/her to a certain level, depending on its strength, his/her awareness or attention and his/her state of mind or mentality ③.

It is important to emphasize the difference between how the observer perceives and interprets a bodily expression, and what impact this expression evokes in the observer. It is expected that the two are related, but there is no information about the nature of this relation or how it evolves and changes over time. One of the goals of this work is to clarify and explain some aspects of this relation to open the possibility of generating an adaptive robot behavior.

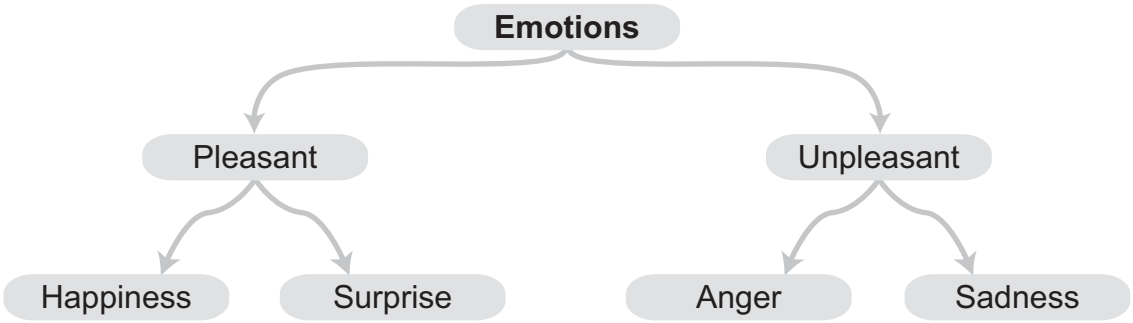


Figure 2.4. The subset of Shaver's classification of emotions [53] used in the categorization of Bodily Expressions.

2.3.1 Classification of bodily expressions

There is a need to classify bodily expressions generated by a robot in order to investigate their effects on the user. For this reason, salient differences among motions should be implemented. During an interaction process, humans go through different affective states, depending on several conditions such as degree of engagement, degree of awareness, and degree of interest among others. It is thus possible to classify every action taking place during an interaction process into the emotional effect that it would have on the observer. I adopted a simplified version of Whissel's wheel of activation-evaluation space described in [64]. I used the fact that we have two primary states for emotions: positive and negative, also known as pleasant and unpleasant emotions. The considered emotions are the following: happiness, surprise, sadness, and anger. In order to categorize these emotions I used a subset of Shaver's classification (see **Figure 2.4**), where happiness and surprise represent pleasant emotions while sadness and anger represent unpleasant emotions [53]. Bodily expressions were classified using one of the specified four emotions and as either pleasant or unpleasant.

2.3.2 Laban features of robot bodily expressions

The humanoid robot ASKA [25] used in this study is shown in **Figure 2.5**. The body has a mobile platform and two arms and is based on the commercial robot

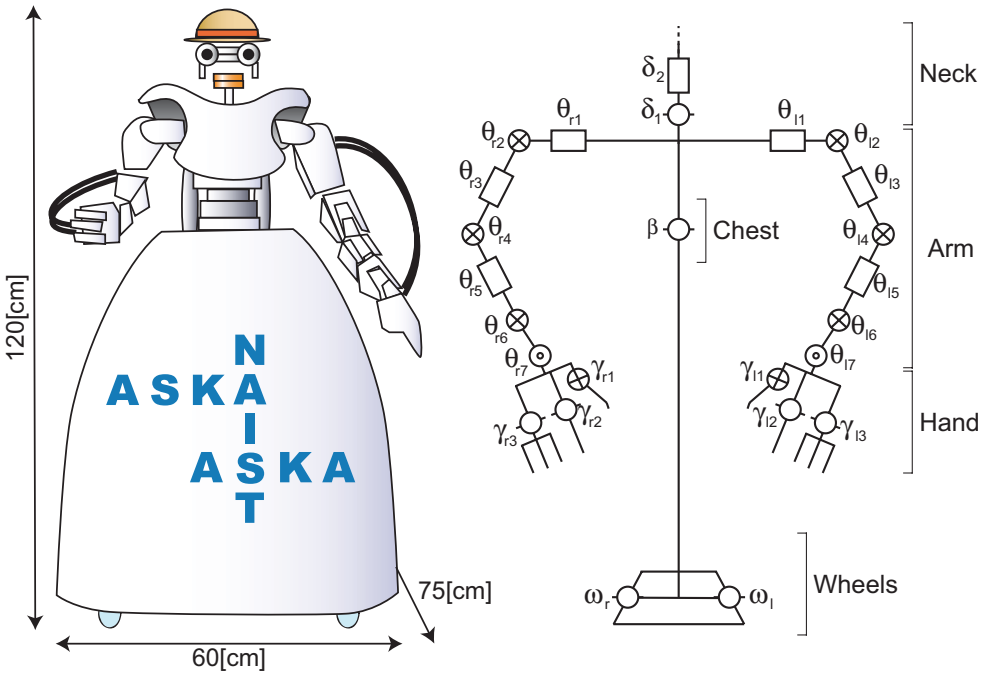


Figure 2.5. Overview of the used robot and its link model.

TMSUK-4[1]. The head is a replica of the Infanoid robot [37]. This humanoid robot with its mobile platform has the advantage of being able to generate relatively fast motions compared to the currently available biped humanoid robots.

Since the pioneering work of Johannson [30] on visual perception of motion, it has been known that humans can perceive a lot of information from body movements including the emotional state of the performer [1, 46]. Recently, there is a growing interest in mathematically modeling emotion-based motion generation for real-world agents such as robots [39] and for virtual agents such as animated characters [3]. To be able to generate bodily expressions that reflect the selected emotions I rely on Laban features of movements [5]. It has been shown in [57] that the qualitative Laban features of *Effort* and *Shape* correlate with the four basic emotions I have considered in section 2.3.1. Based on the mathematical description of Laban features hereafter, it is relatively easy to classify bodily expressions that reflect a certain emotion.

[1] *TMSUK-4* is a trademark of tmsuk Co. Ltd, Kitakyushu.

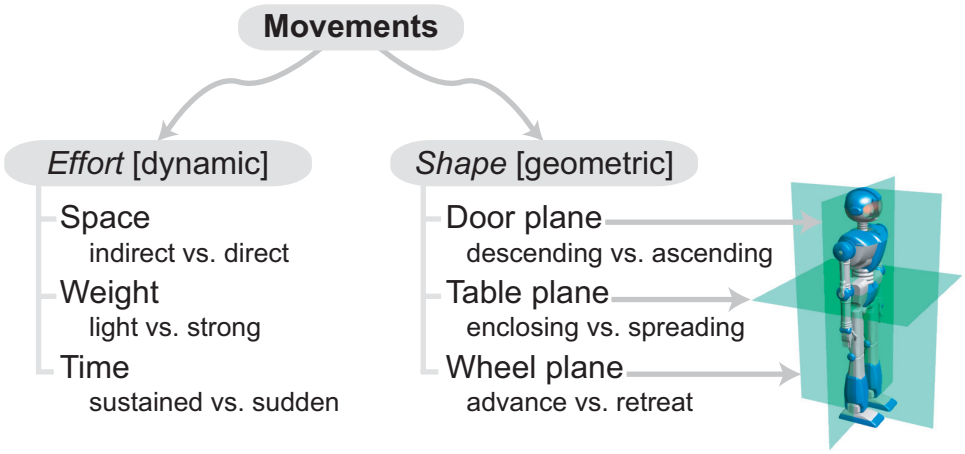


Figure 2.6. Laban features of Shape and Effort have several parameters that describe the geometrical and dynamic aspect of movements.

The mathematical definition of Laban features (*Shape* and *Effort*) using the robot's kinematic and dynamic information is given such that larger values describe fighting movement forms and smaller values describe indulging ones[57]. Bartenieff and Lewis[5] stated that the *Shape* feature describes the geometrical aspect of the movement using three parameters: *table* plane, *door* plane, and *wheel* plane. They also reported that the *Effort* feature describes the dynamic aspect of the movement using three parameters: *weight*, *space*, and *time* (see **Figure 2.6**). The robot's link information which will be used in the features definitions is given in **Figure 2.5**. In order to simplify the mathematical description, a limited number of joint parameters were considered in this definition, namely: the left arm θ_{l1}, the right arm θ_{r1}, the neck δ_1, the face δ_2, the left wheel ω_l, and the right wheel ω_r. The remaining parameters were fixed to default values during movement execution.

Using the diagram shown in **Figure 2.7**, the *table* parameter of feature *Shape* represents the spread of silhouette as seen from above. It is defined as the scaled reciprocal of the summation of mutual distances between the tips of the left and the right hands along with a focus point:

$$Shape_{table} = {}^{s}/(T_{LF} + T_{RF} + T_{LR}) \tag{2.1}$$

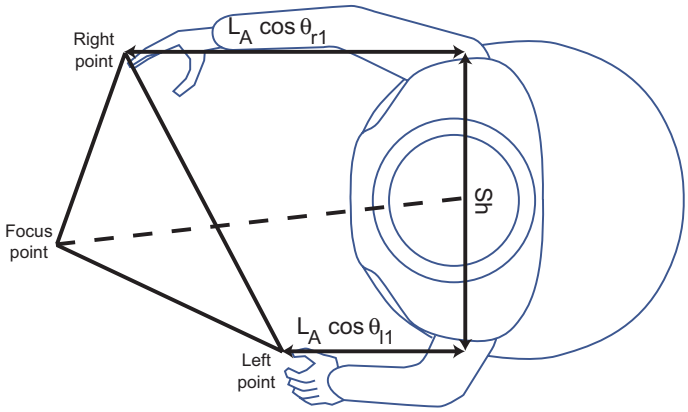


Figure 2.7. Diagram of *table* plane superposed on a top view of the robot.

where,

$$T_{LF} = \begin{bmatrix} (L_F \cos\delta_1 \cos\delta_2 - L_A \cos\theta_{l1})^2 + \\ (Sh/2 - L_F \cos\delta_1 \sin\delta_2)^2 \end{bmatrix}^{1/2},$$

$$T_{RF} = \begin{bmatrix} (L_F \cos\delta_1 \cos\delta_2 - L_A \cos\theta_{r1})^2 + \\ (Sh/2 - L_F \cos\delta_1 \sin\delta_2)^2 \end{bmatrix}^{1/2},$$

$$and \quad T_{LR} = \left[Sh^2 + (L_A \cos\theta_{l1} - L_A \cos\theta_{r1})^2 \right]^{1/2}.$$

The point of focus is set at the fixed distance $L_F = 44[cm]$ in the gaze line of the robot's head. $Sh = 33[cm]$ is the distance between the shoulders, $L_A = 44[cm]$ is the arm's length during movement execution and s is a scaling factor.

The *door* parameter of feature *Shape* represents the spread of the silhouette as seen from the front. It is defined as the weighted sum of the elevation angles of both arms and the head as shown in (2.2). The sine is used to reflect how upward or downward each joint angle is. The weights d_l, d_r, d_n were fixed empirically to 1.

$$Shape_{door} = d_l \sin\theta_{l1} + d_r \sin\theta_{r1} + d_n \sin\delta_{n1} \tag{2.2}$$

The *wheel* parameter of feature *Shape* represents the lengthwise shift of the silhouette in the sagittal plane. It is defined as the weighted sum of the velocities

of the robot and the velocities of the arm extremities as shown in (2.3). Weights were set empirically to -8 for w_t and -1 for w_l and w_r.

$$\begin{aligned} Shape_{wheel} = w_t v_{tr} + w_l L_A \tfrac{d}{dt} \cos\theta_{l1} + \\ w_r L_A \tfrac{d}{dt} \cos\theta_{r1} \end{aligned} \tag{2.3}$$

The *weight* parameter of feature *Effort* represents the strength of the movement. It is defined in (2.4) as the weighed sum of the energies exhibited during movement per unit time at each part of the body. Weights were adjusted with respect of to the saliency of body parts. Relatively large weights $e_{rt} = e_{tr} = 5$ were given to the movement of the trunk and smaller weights were given to the movements of the arms $e_l = e_r = 2$ and the neck $e_n = 1$.

$$\begin{aligned} Effort_{weight} = e_l \dot{\theta}_{l1}^2 + e_r \dot{\theta}_{r1}^2 + e_n \dot{\delta}_{n1}^2 + \\ e_{tr} v_{tr}^2 + e_{rt} v_{rt}^2 \end{aligned} \tag{2.4}$$

where $v_{tr} = \dot{\omega}_l + \dot{\omega}_r$ is the translation velocity and $v_{rt} = \dot{\omega}_l - \dot{\omega}_r$ is the rotation velocity.

The *space* parameter of feature *Effort* represents the degree of conformity in the movement. It is defined in (2.5) as the weighed sum of the directional differences between elevation angles of the arms and the neck as well as the body orientation. Weights are also defined empirically by giving advantage to the arms' bilateral symmetry $s_{lr} = -5$ and body orientation $s_{rt} = -5$ over the other combinations $s_{ln} = s_{rn} = -1$.

$$\begin{aligned} Effort_{space} = s_{rt} \left|\omega_{rt}\right| + s_{lr} \left|\theta_{l1} - \theta_{r1}\right| + \\ s_{ln} \left|\theta_{l1} - \delta_{n1}\right| + s_{rn} \left|\theta_{r1} - \delta_{n1}\right| \end{aligned} \tag{2.5}$$

The *time* parameter of feature *Effort* represents the briskness in the movement execution and covers all the span from sudden to sustained movements. It is defined in (2.6) as the ratio indicating the number of generated commands per time unit.

$$Effort_{time} = \frac{number\ of\ generated\ commands}{time\ span} \tag{2.6}$$

2.3.3 Considered bodily expressions

Although there is no unique solution to this problem, the goal is to be able to generate a representative bodily expression for each of the selected emotions.

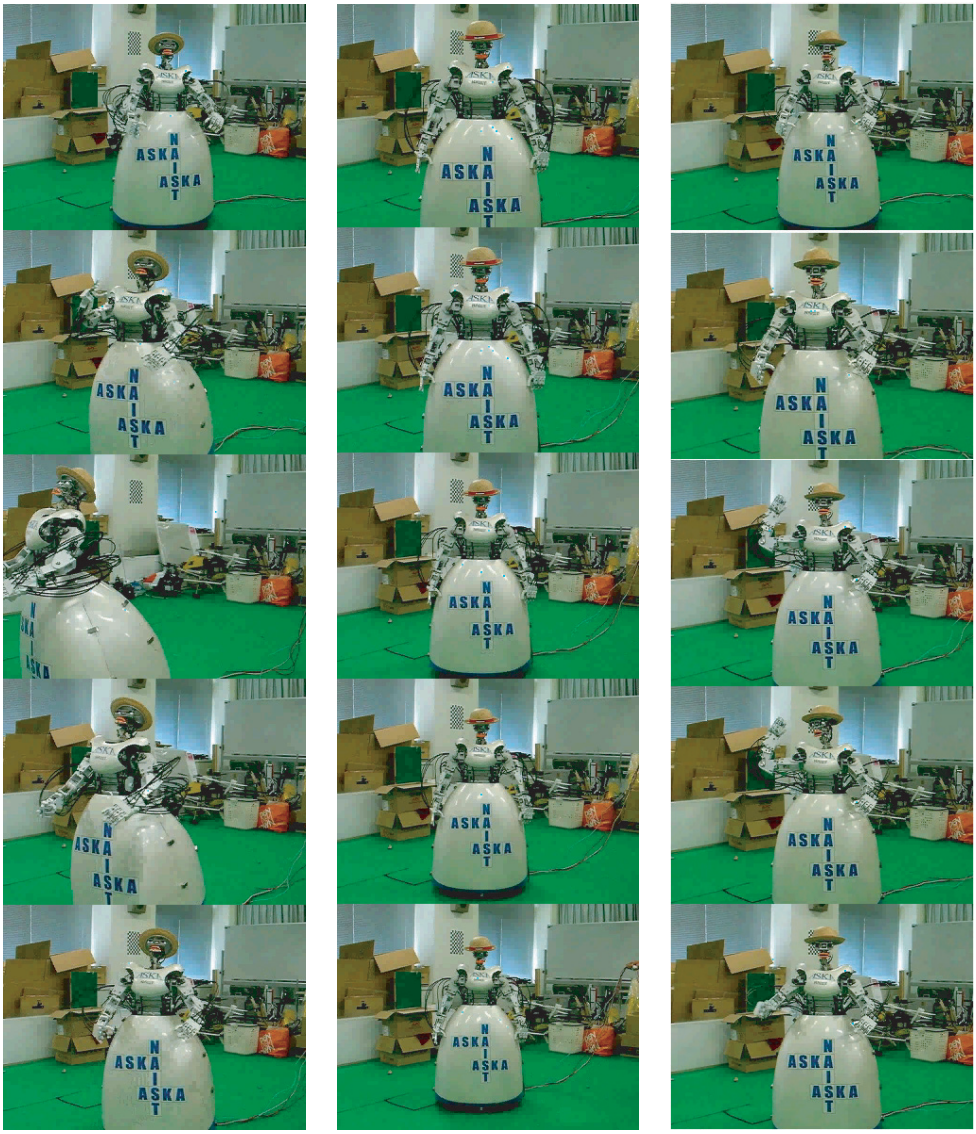
The generated bodily expressions (BE) which are shown in **Figure 2.8** and **Figure 2.9** reflect one of the basic emotions of happiness, surprise, sadness, anger or none. These bodily expressions are the following:

- **BE1 :** The robot raises both arms and turns its body to the left, then to the right, twice. The goal is to show an expression of happiness.
- **BE2 :** The robot raises its right hand and moves it in an arc toward the right side, then goes back to its initial position and lowers its right arm, the goal is to show an expression of no particular emotion.
- **BE3 :** The robot raises both arms and its head, then moves backward for some distance, the goal is to show an expression of amazement or surprise.
- **BE4 :** The robot lowers both arms and its head, and then moves backward at low speed for some distance, the goal is to show an expression of sadness.
- **BE5 :** The robot raises both arms gradually while advancing before stopping, then it lowers and raises its arms progressively for two cycles; the goal is to show an expression of happiness.
- **BE6 :** The robot advances quickly, then goes back and raises its right arm while turning its head a bit to the right. It then lowers its arm and returns its head to the original position; the goal is to show an expression of anger.

The duration of each of these BE was about 14[sec]. Their appropriateness and their expressiveness was tested experimentally using questionnaires (see section 2.4.1).

(a) BE1 (b) BE3 (c) BE5

Figure 2.8. The three bodily expressions generated to represent pleasant expressions (happiness, surprise, and happiness respectively) at different time frames.



(a) BE2 (b) BE4 (c) BE6

Figure 2.9. The three bodily expressions generated to represent unpleasant expressions (neutral, sadness, and anger respectively) at different time frames.

2.3.4 Assessment methods of impression and expressiveness of bodily expressions

There are mainly two types of methods to assess the effects of a particular action on a human. The classic self-reporting approach is widely used, while the assessment of collected physiological data is still an open problem. The first type gives a generally subjective evaluation result; the second type is deemed to be more objective but suffers from inaccuracies. For this case, in order to assess *expressiveness*, I adopted a self-reporting approach and asked the subjects to answer questionnaires. However, in order to assess *impression* the subjects answered questionnaires and their brain activity was recorded.

Subject's answers to questionnaires were summarized and used in order to assess expressiveness. Every subject has to select from: expression of happiness, expression of surprise, expression of sadness, expression of anger, or no meaningful expression. Moreover, the subject had to specify the degree of the expression in a scale of five: 1 for impertinent, 2 for slight, 3 for medium, 4 for strong and 5 for very strong. This selection of the degree of the expression is a redundancy that was meant to confirm the subject's choice and assess the degree of confidence in his/her answer. These answers were then categorized into pleasant or unpleasant expressions using the subset of Shaver's classification shown in **Figure 2.4**.

As for impression assessment, spectral analysis method of electroencephalogram (EEG) data was used. A short EEG segment can be considered as a stationary process, which can be characterized by an autoregressive(AR) model. Let us denote $s(n)$ as a sample of EEG data of N points. I calculate $r_f(n)$ and $r_b(n)$, respectively the forward and backward prediction errors, as follows:

$$r_f(n) = \sum_{k=0}^{p} a(k)\, s(n+p-k) \tag{2.7}$$

$$r_b(n) = \sum_{k=0}^{p} a(k)\, s(n+k), \tag{2.8}$$

where $a(k)$ are the AR parameters and p is the order of the model. The order p is based on the "goodness of fit" criterion. I use the relative error variance (REV)

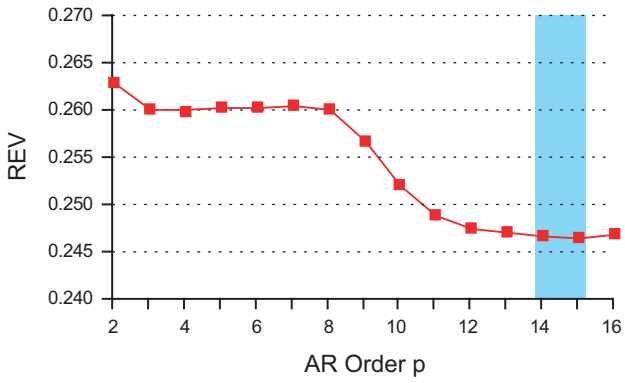


Figure 2.10. Relative error variance (REV) calculated for different orders of AR parameters. Highlighted are the two most optimal choices for the considered data.

criterion[51], defined as:

$$REV\,(p) = \frac{MSE\,(p)}{MSY}, \tag{2.9}$$

$MSE(p)$ is the mean square error or variance of the error process of order p, and MSY is the variance of the total power of the signal sample. The optimal p is the one that minimizes $REV(p)$. In my case, I take $p = 15$ or 14 as shown in **Figure 2.10**.

I apply (2.7) and (2.8) to calculate an approximate estimation of the power spectrum $PS(f)$ of the signal s as follows:

$$PS\,(f) = \frac{V_p\,T}{\left|1 + \sum\limits_{k=0}^{p} a(k)\,e^{-j2\pi fkT}\right|^2} \tag{2.10}$$

$$V_p = \frac{1}{2}\sum_{n=0}^{N+p-1}\left([r_f\,(n)]^2 + [r_b\,(n)]^2\right), \tag{2.11}$$

where V_p is the averaged sum of the forward and backward prediction error energies, and T is the sampling period.

Research in cognitive neuroscience has shown that the low-alpha frequency band is the most reactive band to social cues such as movements [1]. I suppose that this frequency band is going to react in a similar way to robot bodily

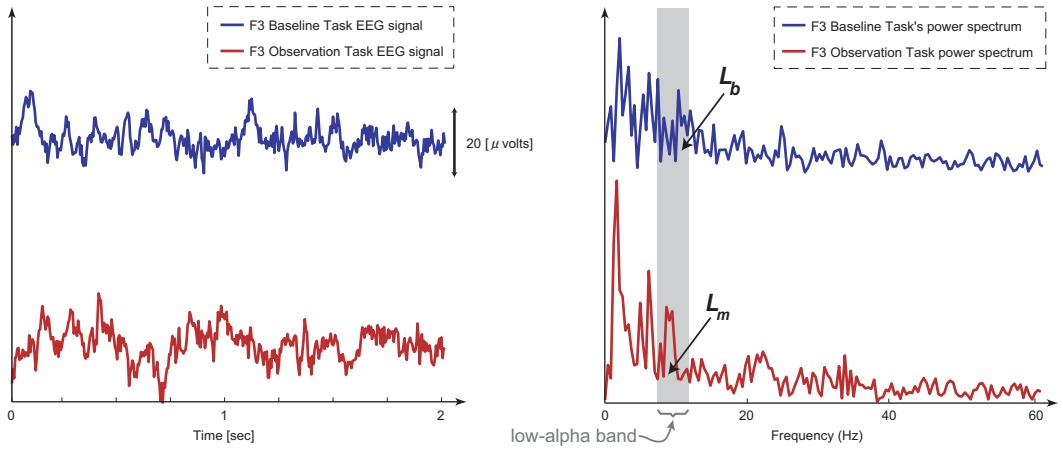


Figure 2.11. An example illustrating the calculation of the power of predefined frequency band, in this case low-alpha band, for a 2[sec] length of raw EEG data taken from electrode placement F3. The graph to the left shows the raw EEG signal for the baseline period and the observation period. The graph to the right shows the power spectrum of the EEG signals, where low-alpha band is highlighted.

expressions [34]. The next step in assessing impression is to observe the amount of change in the power of low-alpha frequency band compared to the whole spectrum. The power L of a band between frequencies a and b is defined by:

$$L\,(a,b) = \frac{\int_a^b PS\,(f)\,df}{\int_0^\infty PS(f)df}. \tag{2.12}$$

Using (2.12), I calculate the power of low-alpha band frequency L_b for the data taken during the baseline period and L_m for the data taken during the period of the execution of a bodily expression. An example illustrating this calculation is shown in **Figure 2.11**, where raw 2 seconds EEG signals collected from placement F3 during the baseline period and the observation period is shown on the left. The power spectrum of these signals is shown to the right, and the low-alpha band is highlighted. A comparison between L_b and L_m will indicate the effect of a particular bodily expression on the user. This is used as the main evaluation criterion for impression.

Table 2.1. Users' evaluations of the expression of each generated robot bodily expression (BE)

BEs	Pleasant	Unpleasant	Neither
BE1	**100**%	0 %	0 %
BE2	6 %	35 %	**59**%
BE3	**94**%	6 %	0 %
BE4	0 %	**94**%	6 %
BE5	**65**%	12 %	23 %
BE6	0 %	**82**%	18 %

2.4. Experimental study

2.4.1 Expressiveness of robot bodily expressions

The goal of this experiment is to evaluate the expressiveness of each generated robot bodily expression. Since this quality is highly subjective, the self-reporting approach was used.

Subjects

Seventeen (17) participants (two females and fifteen males aged between 20 and 50 years old) volunteered to take part in this experiment. They were either students or faculty members. They were all familiar with robots and had previous experiences of dealing with robots similar to the one used in the experiment.

Procedure

Every subject was shown a total of six bodily expressions, which were described in 2.3.3. The execution of each of the bodily expressions by the humanoid robot ASKA lasted 14 seconds. After observing each robot bodily expression, enough time was given to the subject to answer two questions about the expressiveness of that robot bodily expression, and one more question about their impression after

Table 2.2. Users' evaluations of their impressions after observing each robot bodily expression (BE)

BEs	Pleased	Unpleased	Neither
BE1	**65**%	35 %	0%
BE2	30 %	**70**%	0%
BE3	**68**%	32 %	0%
BE4	19 %	**81**%	0%
BE5	**100**%	0 %	0%
BE6	47 %	53 %	0%

the observation. These answers were then summarized to assess the expressiveness as explained in 2.3.4.

Results

Table 2.1 shows the results about bodily expressions after categorization into *pleasant expression*, *unpleasant expression*, or *neither*, clearly indicating the expressiveness as evaluated by the observers. The result about impressions is presented in **Table 2.2** after categorizing the answers into *pleased* or *unpleased*.

These results demonstrate the existence of a strong correlation between the expressiveness of the robot bodily expressions as seen by the subjects and the target expression when these bodily expressions were generated (see 2.3.3). BE1, which was created to express happiness, was classified as having a 100% pleasant expression. BE2, which was created to express a neutral emotion, was classified by 59% as neither pleasant nor unpleasant, and by 35% as unpleasant, suggesting that neutral bodily expressions tend to have a negative connotation. BE3, which was created to express surprise, was classified by 94% as a pleasant expression. BE4, which was generated to express sadness, was classified by 94% as being an unpleasant expression. Similarly, BE6 which was created to express anger was also classified by 82% as an unpleasant expression. The special case of BE5 was classified to a great extent as a pleasant expression by up to 65%. However, 23%

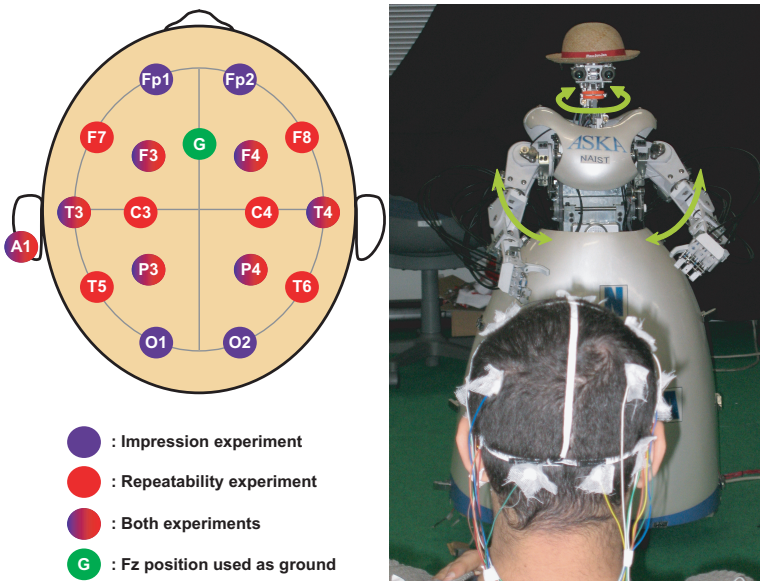


Figure 2.12. The experimental setup where brain activity was measured according to the 10-20 international standard[29].

said it did not express anything in particular and 12% claimed it was unpleasant.

The expensiveness of the generated BEs is confirmed to be in accordance with the target expressions for which they were created. BEs generated to express happiness and surprise expressions were classified as pleasant, and the BEs generated to express sadness and anger expressions were classified as unpleasant. Among the generated BEs I choose one that is representative of each category in order to use it in the evaluation of its impressions on the observer.

2.4.2 Impressions of robot bodily expressions

The goal of this experiment is to evaluate the impression on the observer of the generated motions using a hybrid approach that combines the results of self-reporting and analysis of brain activity.

Subjects

Seven (7) participants (one female and six males, 23∼43 years old) volunteered to take part in this experiment. They were all students or faculty members, and only two of them had the experience of using electroencephalography to collect brain signals. Before starting the experiment each participant was fitted with the electrodes while allowing him/her to spend more than 20 minutes reading books of interest to become familiarized to the electrodes' presence.

Procedure

During each session, 12 EEG channels (using sintered Ag/AgCl electrodes) were recorded by the 5200 Series DigitalAmp System[2]. The recording was performed from 10 placements, namely: Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, O2 according to the international 10-20 standard (see **Figure 2.12**). The placement Fz was used as the ground, and the signal from the left ear placement A1 was used as the reference signal. The contact impedance between all electrodes and the skin was kept below $5[k\Omega]$. The subjects were shown a total of six motions lasting 14 seconds each by the humanoid robot ASKA while their brain activity was recorded with 16-bit quantization at a sampling frequency of $200[Hz]$.

The subjects were asked to relax as much as possible and not to think of anything in particular when recording the baseline period, which lasted for 14[sec]. They were also told that they would be asked about the robot's movements and that they had to watch carefully when the robot was moving. This is important because I need to make sure that the subjects attended to the task. After the observation of each bodily expression, the subjects described their impression in their own words. Having no constraint to express themselves, the subjects gave more details about their impression. These answers were used in categorizing the impressions into pleased or unpleased based on Shaver's original classification of emotions [53].

[2]The 5200 Series DigitalAmp System is a trademark of NF Corporation, Yokohama.

Table 2.3. Significant change in low-alpha power according to observed motion categories at every electrode and for each subject. (+: significant change $p < .05$; −: no significant change)

Subject	Category	Electrodes									
		Fp1	Fp2	F3	F4	T3	T4	P3	P4	O1	O2
1	Pleasant	-	-	-	+	-	-	-	-	+	-
	Unpleasant	-	+	-	+	+	-	-	-	+	-
2	Pleasant	-	-	+	-	-	-	-	-	-	-
	Unpleasant	-	-	-	-	-	-	+	+	-	-
3	Pleasant	-	-	-	-	+	+	-	+	-	-
	Unpleasant	-	-	-	-	-	-	-	-	-	-
4	Pleasant	-	-	-	-	-	+	+	+	-	-
	Unpleasant	-	-	-	-	-	-	+	+	+	-
5	Pleasant	-	-	-	-	-	+	-	-	-	-
	Unpleasant	-	-	-	-	+	+	+	+	-	-
6	Pleasant	+	+	-	-	+	+	-	-	-	-
	Unpleasant	-	-	-	-	+	+	-	-	-	-
7	Pleasant	-	-	+	-	+	+	-	-	-	+
	Unpleasant	-	-	+	-	-	-	-	+	-	-

Results

Table 2.2 shows the self-reporting result about the subjects' impressions after observing every robot bodily expression. There is a strong correlation between these results and the expression results, reported previously in 2.4.1, with a coincidence level of 71%. For example, BE4 impression was considered to be unpleasant by up to 81% and its expressiveness was considered unpleasant by 94%. This is also the case for BE1 where its impression of being pleasant is 65%, and it expression of being pleasant is 100%. The same could be said for BE3, with a pleasant impression of 68% and a pleasant expression of 94%. The case of BE6 is different from the previous ones. While its expression was considered unpleasant by 82%, its impression shows the small rate of 53% for being unpleasant and 47% for being pleasant. It is still inclined to the unpleasant side. However, its pleasant effect cannot be explained knowing that this bodily expression was created to express anger. The last case of BE2 shows a big difference between its 59% neutral expression and its 70% unpleasant impression. This suggests that bodily expressions with a neutral expression can be perceived negatively and generate an unpleasant impression.

The analysis of EEG data using the method described in 2.3.4 allowed the calculation of the power L_m of low-alpha band in each electrode channel and for each bodily expression. It also allowed the calculation of the power L_b of the same band for the baseline period. Comparing L_m and L_b reveals the effect of observing a bodily expression in the change in the power of low-alpha band for each electrode channel. **Table 2.3** summarizes the results of this change in power, where only statistically significant change is indicated with the symbol +. It can be seen that significant effect is mostly present at locations T3 and T4, then at P3 and P4, and finally at F3 and F4. Knowing that these positions are located above the superior temporal sulcus (STS) and above some specific parts of the prefrontal cortex (PFC) confirms previous research findings about the activation of STS in the perception of social cues [8, 1], and the activation of the mirror neurons located in the PFC during learning and imitation tasks

[49]. Some reaction can also be seen at other locations, for instance O1 and Fp2 for subject 1, O2 for subject 7, Fp1 and Fp2 for subject 6. The reaction at locations Fp1 and Fp2 are thought to be the result of blinking activity during the recording process, since these electrode positions are the closest to the eyes. It is important to assert that no preprocessing was done to avoid data with eye blinking artifacts. This approach was adopted because the goal is to conduct this investigation in natural conditions, where blinking activity is normal and unavoidable. The reaction at locations O1 and O2 could be explained by the fact that during the vision process the visual cortex gets activated and this activation is usually captured at locations O1 and O2.

Nevertheless, the reactive locations were not always the same among different observers, suggesting high individual differences. A generalization cannot be derived at this point about the reaction of brain locations according to the category of the bodily expression that is being observed. However, the presence of a reaction is confirmed and another approach is necessary to achieve a more comprehensive result. On the other hand, there is a need to assess the repeatability of similar reactions from the same observer when he/she is shown the same bodily expression.

2.4.3 Repeatability of reaction in brain activity

The goal of this experiment is to confirm that the results obtained in the impression experiment (see 2.4.2) are consistent over time. In other words, to make sure that brain reaction does happen all the time and in the same way if a subject observes the same bodily expression several times.

Subject

One (1) student (male, 32 years old) volunteered to take part in this experiment. The subject was fitted with an electro-cap[3] and was given about 30 minutes to familiarize him to the presence of the cap.

[3]Electro-Cap is a trademark of Electro-Cap International Inc., USA.

Procedure

The subject participated in ten recording sessions. In each session, he was shown two bodily expressions, one for each category of bodily expressions, executed by the humanoid robot ASKA. Showing only representative bodily expressions is sufficient since the goal is to confirm the repeatability of brain reaction. Each bodily expression lasted for 14[sec], and the baseline period was recorded during the 14[sec] before the execution of each bodily expression. BE1 was chosen as a representative of pleasant bodily expressions, and BE4 was chosen as representative of unpleasant bodily expressions. On one hand, BE1 was chosen because its expressiveness evaluation as pleasant (100%) was the highest among all the bodily expressions. Its impression evaluation (65%) was high enough to make sure it will have the desired effect on the observer, even though its impression was evaluated as the lowest among all the pleasant bodily expressions. In this case, the advantage was given to the expressiveness evaluation over the impression evaluation. On the other hand, BE4 was chosen because its expressiveness evaluation as unpleasant (94%) was the highest among all the bodily expressions. Its impression evaluation (81%) was also the highest among all the bodily expressions, making it the perfect candidate to represent unpleasant bodily expressions.

The recording was performed from 12 placements of an electro-cap, namely: F3, F4, C3, C4, P3, P4, F7, F8, T3, T4, T5, and T6 (see **Figure 2.12**), using the Polymate AP1132[4] EEG recording device. The sampling frequency was set to 200[Hz] and the impedance was kept below $5k\Omega$. As a result of the experiment in 2.4.2, the placements Fp1, Fp2, O1, and O2 were omitted, since they were shown to have an insignificant reaction. On the other hand, new placements were introduced, namely: F7, F8, C3, C4, T5, and T6, in order to get a more detailed coverage of the sensory-motor cortex as well as uncovered regions in the previous experiment.

[4]Polymate AP1132 was developed by Digitex Lab. Co. Ltd, and is commercialized by TEAC Corporation, Tokyo.

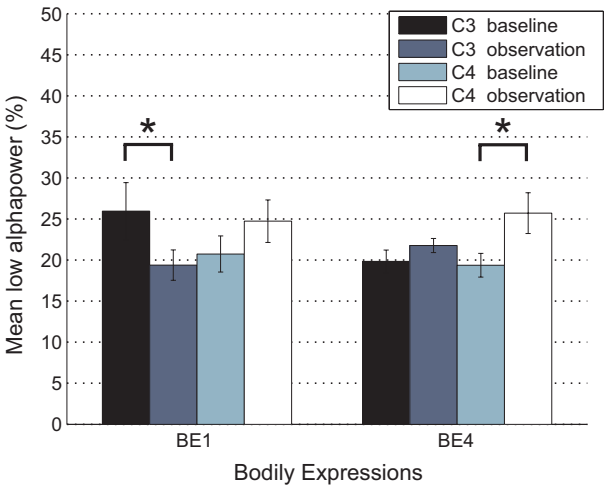


Figure 2.13. Mean alpha power calculated for the ten trials and for electrodes C3 and C4. ($\star : p < .05$)

During the recording of the baseline period, the subject was asked to relax as much as possible and to think of nothing in particular. To confirm that the subject attended to the task, he was told that he would be asked about the robot's movements and that he had to observe carefully. After the observation of each bodily expression, the subject was asked to explain the difference between the recently observed bodily expression and the one just before it.

Results

Analysis of the collected data, using the method described in 2.3.4, resulted in identifying the electrode channels of placements C3 and C4 as the most reacting for this subject. **Figure 2.13** shows the mean power of low-alpha frequency band calculated from the 10 trials for the electrode placements C3 and C4: where C3 reacted significantly to the pleasant bodily expression and C4 reacted significantly to the unpleasant bodily expression.

Figure 2.14 shows the overall result for the two bodily expressions by averaging the power change for all the electrodes over the 10 trials. The difference in means is significant between BE1 and BE4. Since BE1 is representative of

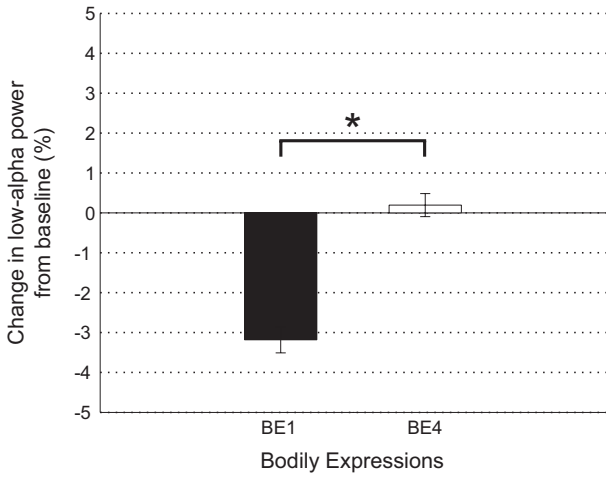


Figure 2.14. Change in alpha power from the baseline using all electrodes, for each of the considered unpleasant and pleasant motions. ($\star : p < .05$)

pleasant bodily expressions and BE4 is representative of unpleasant ones, this result suggests an overall significant decrease in the power of low-alpha frequency band for pleasant motions, and a significant increase in power of low-alpha frequency band for unpleasant ones. This confirms that the change in low-alpha power happens every time the observer watches a bodily expression, and that this change is inversely proportional to the category of the observed bodily expression.

2.4.4 Discussion

The results presented in **Table 2.1** confirmed the appropriateness of the expressiveness of the generated bodily expressions used in the experimental study. These show the unpleasant bodily expressions were classified as such, and pleasant bodily expressions were also classified as such. During every experiment, the order of which the bodily expressions were shown was random so as not to allow the user to predict the nature of the next bodily expression. Although the subjects were not told anything about the bodily expressions, their answers agreed with the hypothesis. This implies that people tend to see bodily expressions in similar ways, which facilitates their interpretations and use in interactions. It can be

thought that there exist a basic knowledge that is shared among us to allow proper interpretation of similar expressions, although this knowledge is highly affected by the environment factors of culture and ethnics. The bodily expressions were treated by the users as if they were performed by a human even though it was the robot ASKA which actually performed them. It would be interesting to compare the interpretation of the same bodily expressions executed by both humans and robots to evaluate the existence of interpretation differences.

On the other hand, **Table 2.2** correlates to a high extent with the results of **Table 2.1**. Here I can infer that observing a pleasant bodily expression will result in a pleasant impression on the observer and vice versa. This means that the observer is affected by what he sees even though the actor is just a robot. This effect on the observer is shown to be present in his/her brain activity with the result presented in 2.4.2. Although, a generalization could not be concluded from these results, the presence of a reaction in brain activity was proven. It is important to acknowledge that the most reactive electrode position were F3, F4, T3, T4, P3 and P4, which are located above the STS and PFC. This supports the claims that STS and mirror neurons get activated during the perception of social cues and the observation of movements [1, 8], and that this can be used effectively in the implementation of Brain-Machine Interfaces [42, 63].

Finally, it was necessary to confirm the repeatability or the reproducibility of the same reaction in similar conditions. The results showed that the low-alpha power level over all brain activity decreased or increased in accordance to the category of the observed bodily expression. Particularly the most significant reaction was present at electrode positions C3 and C4. These positions are close to the premotor and motor cortices. Due to the low spatial resolution of EEG, it is difficult to assert precisely which part of the brain is responsible for these reactions. However, current research findings confirm that the STS has an important role in the interpretation of social cues [1], and that mirror neurons are important during learning and imitation tasks [49].

2.5. Conclusion

In this chapter, I investigated the relation between bodily expressions and their impressions on an observer. I showed that this problem is necessary to advance the current status of human-robot interaction. The necessity comes from the fact that tacit information has been neglected in previous research work. For this reason, I showed that this study is related to and can benefit from the achievements of two fields: brain-machine interfaces and cognitive science. I started by generating six bodily expressions, then I classified them into two categories according to their expressiveness, namely: pleasant and unpleasant. Their expressiveness was confirmed statistically with the result of a self-reporting experiment where a number of volunteers answered questionnaires about the bodily expressions. Afterwards, I conducted an experiment to assess the impressions on the observers while watching the considered bodily expressions by measuring their brain activities using electroencephalogram (EEG). The method adopted for spectral analysis revealed a correlation between the power level of low-alpha band (8-11$[Hz]$) and the category of the observed bodily expression. The reproducibility or repeatability of this band's reaction was confirmed with a third experiment where a subject observed candidate bodily expressions for each category.

These results have opened the opportunity to utilize the change in the power level of low-alpha frequency band to examine the capacity of a humanoid robot in activating the social perception system in a human observer. A challenging problem that rises from this result is about the degree to which such reaction appears when observing robots with different human-like physical and behavioral characteristics. The understanding of which robot properties are necessary or sufficient to activate the social perception system in an observer is of particular interest. Another important direction is to define computational methods which can assess and recognize the impression category from the observer's brain activity, and this is the issue considered in the next chapter. Potential applications include customized interface adaptation to the user, interface evaluation, or simple user monitoring.

Chapter 3

Recognition of Impression from Brain Activity

3.1. Introduction

Extracting useful information and recognizing specific mental states from brain activity is a challenging research problem. This is mainly due to the high presence of noise in the treated data, especially EEG measurements. This problem is generally viewed as the problem of reduction of the high-dimensionality present in the EEG data. Research on the development of Brain-Machine Interfaces (BMI) has grown toward the classification of user's intents from EEG patterns [66]. A major problem with BMI implementations is the adaptive training of classifiers. There are mainly two reasons for using adaptive training processes. The first one is that the subjects cannot maintain exactly the same mental states for a long time. EEG patterns recorded at different times vary to a big extent even if the subject performs the same task. The second reason is that it is hard to provide sufficiently large data sets for the training of classifiers. This can only be solved by adopting an adaptive and unsupervised training policy [58].

Unsupervised training methods are important, as their training does not require prior knowledge, for instance labeled data segments. They can also adapt to the variability of EEG patterns by adopting a continuous learning

approach. In the following sections I will start by an overview of existing methods for data exploration. Then I will explain about human bodily expressions in contrast to robot bodily expressions. After that, I will introduce an unsupervised classification method that can help in the recognition of the impressions on an observer of bodily expressions of robots and humans alike. Finally, an improved version of this method is explained and shown to be more robust against noise, variance in EEG patterns and individual differences.

3.2. Related research

Several graphical means have been proposed to visualize high-dimensional data item directly, by linking each dimension to an aspect of the visualization and integrating the result into a single figure [59, 27]. These methods can be used to visualize any kind of data by viewing the data items themselves or vectors formed of some descriptors of the data set [62].

The simplest way to visualize data sets is to plot a descriptor of each item, i.e. a two dimensional graph in which the dimensions are enumerated on the x-axis and the corresponding values on the y-axis (the histogram is a good example of such a representation). An alternative is a scatter plot where two original dimensions of the data are chosen to be portrayed as the location of an icon, and the rest of the dimensions are depicted as properties of the icon (the pie diagram is a good example of such a representation).

The major drawback of these methods when dealing with high-dimensional data is that they do not reduce the amount of data. If the data set is large, then the display consisting of all data items will be incomprehensible. Fortunately, there exist other kinds of methods which try to reduce the amount of data using some sophisticated calculations. Representative examples of these methods will be described in the following.

3.2.1 Clustering methods

The goal of clustering is to reduce the amount of data by grouping or categorizing similar data items together. The main motivation for using these methods is to provide automated tools to help making taxonomies [4, 18, 28, 55, 61]. These methods may also be used to minimize the effects of human errors during the process. Clustering methods can be divided into two basic types: hierarchical and partitional clustering. Within each type there exist several subtypes and different algorithms for cluster extraction.

Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. The clustering methods differ in the rule by which it is decided which two small clusters to merge or which large cluster to split. The final result of the algorithm is a tree of clusters which shows how they are related to each other.

Partitional clustering attempts to directly decompose the data into a set of disjoint clusters. The criterion function that needs to be minimized may emphasize local structure of the data, for instance by assigning clusters to peaks in the probability density function. The global criteria involve minimizing dissimilarity between items belonging to the same cluster, while maximizing dissimilarity between different clusters.

A commonly used partitional clustering method is the *K-means clustering* [40]. In this method, the criterion function is the average squared distance of data items $\mathbf{x}_k$ from their nearest centroids,

$$E_K = \sum_k \left\| \mathbf{x}_k - \mathbf{m}_{c(\mathbf{x}_k)} \right\|^2, \tag{3.1}$$

where $c(\mathbf{x}_k)$ is the index of the centroid that is the closest to $\mathbf{x}_k$. One of the algorithms for minimizing the cost function begins by initializing a set of K cluster centroids denoted by $\mathbf{m}_i, i = 1, ..., K$. The positions of $\mathbf{m}_i$ are then adjusted iteratively by first assigning the data samples to the nearest clusters and then re-computing the centroids. The iteration is stopped when E_K does not

change significantly anymore. Equation (3.1) is also used in a related method called *vector quantization* [12, 14, 41]. This method's goal is to minimize the average quantization error, which is the distance between a sample $\mathbf{x}$ and its representation $\mathbf{m}_{c(x)}$.

Two major problems exist with the clustering methods. First, the interpretation of the clusters may be difficult. Most algorithms prefer certain cluster shapes, so they assign data to clusters of such shapes even if there were no clusters in the data. It is essential to analyze whether the data set exhibits a clustering tendency and the result of cluster analysis needs to be validated as well [27]. The second problem is that the choice of the number of clusters is critical; different kinds of clusters may emerge when K is changed. Proper initialization of the cluster centroids may also be crucial; some clusters may even be left empty if their centroids were initialized far from the distribution of data.

Clustering can be used to reduce the amount of data and to induce categorization. However, the generated categories would need another simplification since their representation is still high-dimensional.

3.2.2 Projection methods

Projection methods are meant to reduce the dimensionality in the data. Their goal is to represent the input data items in a lower-dimensional space in such a way that certain properties of the structure of the data set are preserved as best as possible.

Linear projection methods use linear transformations to project the input data into a lower-dimensionality space. This category of methods include the *principal component analysis* (PCA) and the *projection pursuit* method.

PCA [22] can be used to display the data as a linear projection on a subspace of the original data space that best preserves the variance in the data. It is a standard method in data analysis; it is well understood and effective algorithms exist to compute the projection.

In projection pursuit [10, 11], the data is projected linearly with a projection that reveals as much as possible the non-normally distributed structure of the data set. This is done by assigning a numerical "interestingness" index to each possible projection and by maximizing this index. The definition of interestingness is based on how much the projected data deviates from normally distributed data in the main body of its distribution.

Nonlinear projection methods give a representation for each data point in the lower-dimensional space and try to optimize these representations so that the distances between them would be as similar as possible to the original distances of the corresponding data items. The methods differ in how the different distances are weighted and how the representations are optimized.

Multidimensional scaling (MDS) [60, 67, 32] refers to a group of methods that is widely used especially in behavioral, econometric, and social sciences to analyze subjective evaluations of pairwise similarities of entities, such as commercial products in a market survey. The starting point of MDS is a matrix consisting of the pairwise dissimilarities of the entities. It is most often used to create a space where the entities can be represented as vectors, based on some evaluation of the dissimilarities of the entities. Often linear projection onto a subspace obtained with PCA is used. The key idea is to approximate, using a nonlinear projection method, the original set of dissimilarities with distances corresponding to a configuration of points in an Euclidean space.

Principal curves [19] are a generalization of PCA to form nonlinear curves. They are smooth curves defined with the property that each point of the curve is the average of all data points that project to it, i.e., for which that point is the closest point on the curve. Intuitively speaking, the curves pass through the "center" of the data set.

3.2.3 Hybrid methods

These methods are a special case that tries to combine the properties of the previously introduced two categories. They can be used at the same time to reduce the amount of data by clustering and to reduce their dimensionality by a nonlinear mapping onto a lower-dimensional space. This category of hybrid methods includes the self-organizing maps (SOM) [35] and its variants. A detailed description of SOM and one particular variant will be given in the following sections.

Before that, a description of an additional data set used in my study is given in the next section. This data set will extend the conclusion drawn in the previous chapter that brain activity reacts to robot bodily expressions. A comparison between brain reaction to robots and humans would cover the span of possibilities considered for physical characteristics that activates the social perception circuits in the observer's brain.

3.3. Human bodily expressions

There is little knowledge of the difference in the effect left on a person when observing humans and when observing robots [26, 1, 52]. Most of the literature report either cases and there is, to my knowledge, no previous work that tried to draw on the parallels between the two cases. Therefore, it is interesting to investigate similarities and differences in the recognition of the category of the impression left on the observer for both cases.

In order to collect brain activity data when subjects are observing human bodily expressions, I conducted an experiment similar to the one for the expressiveness of robot bodily expressions (see 2.4.2). The goal is to evaluate the impression on the observer of the bodily expressions described in 2.3.3.

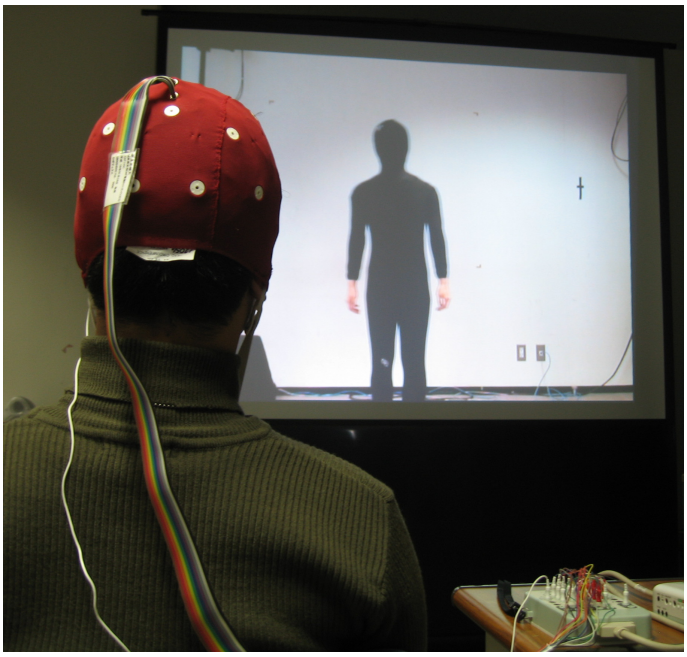


Figure 3.1. Overview of the experiment where a subject observes human bodily expressions.

3.3.1 Data acquisition

Subjects

Seven (7) participants (males aged between 21 and 23) volunteered to take part in this experiment. They were student at the Graduate School of Information Science. They were all familiar with the experiment since they did participate in the previous one about impressions of robot bodily expressions. Their brain activity was collected with an EEG measurement device and they were familiarized with the presence of the electrodes by letting them spend about 20 minutes reading books or surfing the Internet. An overview of the experiment where the subjects observed human bodily expressions is shown in **Figure 3.1**.

(a) BE1 (b) BE3 (c) BE5

Figure 3.2. The three human bodily expressions generated to represent pleasant expressions (happiness, surprise, and happiness respectively) at successive time frames.

(a) BE2 (b) BE4 (c) BE6

Figure 3.3. The three human bodily expressions generated to represent unpleasant expressions (neutral, sadness, and anger respectively) at successive time frames.

Procedure

The human bodily expressions that were shown to the subjects were prepared beforehand. A volunteer in a black tight suit performed these bodily expressions. He had a black cover on his head to make sure that no facial expressions get to the observer (see **Figure 3.2** and **Figure 3.3**). These bodily expressions were a reproduction of the bodily expressions described in 2.3.3 and were recorded on HDTV[1] video tapes. During the experiment, the bodily expressions were projected on a display big enough to ensure that the projected images of the human performer was as close as possible to real size.

Similar to the experiment in section 2.4.2, the recording was obtained from 12 electrode locations, namely: F3, F4, F7, F8, C3, C4, P3, P4, T3, T4, T5, T6, using the Polymate AP1132 EEG recording device. The sampling frequency was $200[Hz]$ and the impedance was kept below $5[k\Omega]$. The subjects were shown a total of six bodily expressions lasting $10[sec]$ each. During the recording of the baseline period, the subjects were asked to relax as much as possible and to think of nothing in particular. To help them achieve this state of mind, they were shown for $10[sec]$ the empty space of the room where the recording of the bodily expressions was performed. Moreover, in order to confirm that the subjects attended to the task they were told that they would be asked about the bodily expressions and that they had to observe carefully. After each observation, they were asked to explain the difference between the bodily expressions they just observed and the previous ones.

The collected data were to be used, along with the data for robot bodily expressions, in the recognition of the impressions using computational methods described in the next sections.

[1]HDTV stands for High Definition television a.k.a High-Vision which allows the recording of a high resolution video stream.

3.4. Recognition with Self-Organizing Maps

3.4.1 Self-Organizing Maps (SOM)

The Self-Organizing Map or Kohonen Network is a neural network algorithm well suited to represent and generalize input data with an underlying structure that is not easily grasped [35]. Its main characteristic is that it can be used at the same time to reduce the amount of data by clustering and to project the non-linear data into a lower-dimensional, usually 2D, map.

The unsupervised and competitive learning process tries to represent the high-dimensional data rather than classify them. It is an adaptive process where the neurons of the map gradually try to become sensitive to different categories of input data [2]. Different neurons specialize in representing different types of inputs. This specialization is enforced by competition among neurons. When a d-dimensional input vector $\mathbf{x} = (x_1, \ldots, x_d)$ arrives, the neuron $\mathbf{m}_c$ that best represents it wins the competition and is adapted to represent it even better:

$$\mathbf{m}_c = \arg\min_i \left(D\left(\mathbf{x}, \mathbf{m}_i\right)\right), \tag{3.2}$$

where D is some measure of similarity or distance between the input vector $\mathbf{x}$ and a neuron $\mathbf{m}$. The winner neuron's neighbors also adapt and gradually specialize to represent similar inputs. Both the metric relation and the probability density of the input data is approximated during this process, allowing the classification of newly collected data. Each neuron in the map is represented by a d-dimensional vector $\mathbf{m} = (m_1, \ldots, m_d)$ called a prototype vector (a.k.a. weight vector or codebook vector). During the learning process, an arbitrary prototype vector, $\mathbf{m}_j$, is updated at time t by:

$$\Delta\mathbf{m}_j\left(t\right) = \alpha\left(t\right) h_{cj}\left(t\right)\left[\mathbf{x}\left(t\right) - \mathbf{m}_j\left(t\right)\right], \tag{3.3}$$

where $\alpha\left(t\right)$ is a decreasing learning rate factor, and $h_{cj}\left(t\right)$ is a decreasing neighborhood function between $\mathbf{m}_c$, the prototype vector winning the competition, and the prototype vector $\mathbf{m}_j$.

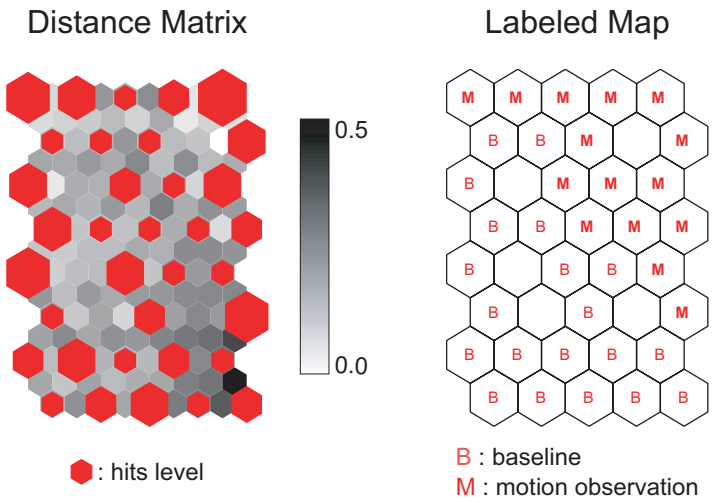


Figure 3.4. Example of the distance matrix and the labeled map of a SOM trained with data representing two categories, one recorded during the baseline, and the other when a subject was observing BE2. The labeling of the map using representative data revealed the clusters where the input data of each category is mapped.

Once the learning process has terminated, a labeling of the map can be done by using a relatively small set of data samples previously labeled. An example of a trained and labeled SOM is shown in **Figure 3.4**. The resulting map could be used to monitor the topographic patterns related to specific events [31], or it could be used in the recognition of newly obtained data. The similarity between a d-dimensional feature vector $\mathbf{x}$ and a prototype vector $\mathbf{m}$ in the learned map is calculated using a variant of Minkowski metric defined by:

$$D_P(\mathbf{x}, \mathbf{m}) = \left(\sum_{j=1}^{d} w_j \left| x_j - m_j \right|^P \right)^{\frac{1}{P}}, \tag{3.4}$$

where w_j is a weighting factor which can be used to give preference for certain features over others ($0 \leq w_j \leq 1$). This proved helpful and important in the semi-assisted learning of the data structure that was necessary for the EEG data. When $P = 1$, equation (3.4) becomes a weighted Manhattan metric, and when $P = 2$, it becomes a weighted Euclidean distance. The latter is selected as a similarity measure to use in SOMs. The usual approach in using SOM starts by preprocessing the selected data. Then, a feature extraction method is specified and used. After that, the map is calculated using competitive learning. Finally, the resulting map is used for recognition. It is important to note that in practical applications the selection and preprocessing of data is extremely important, because unsupervised methods only illustrate the structure in the data set; and the structure is highly influenced by the features chosen to represent data items [33]. In the following section, I will show how I used SOM in the recognition of bodily expressions executed by a humanoid robot and by a human.

3.4.2 Feature selection

In the recognition of the impressions of robot bodily expressions I use the same data as in 2.4.2. The data consists of all signals collected at a sampling rate of 200$[Hz]$ from 10 EEG placements for 2×14 seconds and for each one of the seven subjects. For the recognition of human bodily expressions I use the data described in 3.3. This data consists of all signals collected at a sampling rate of

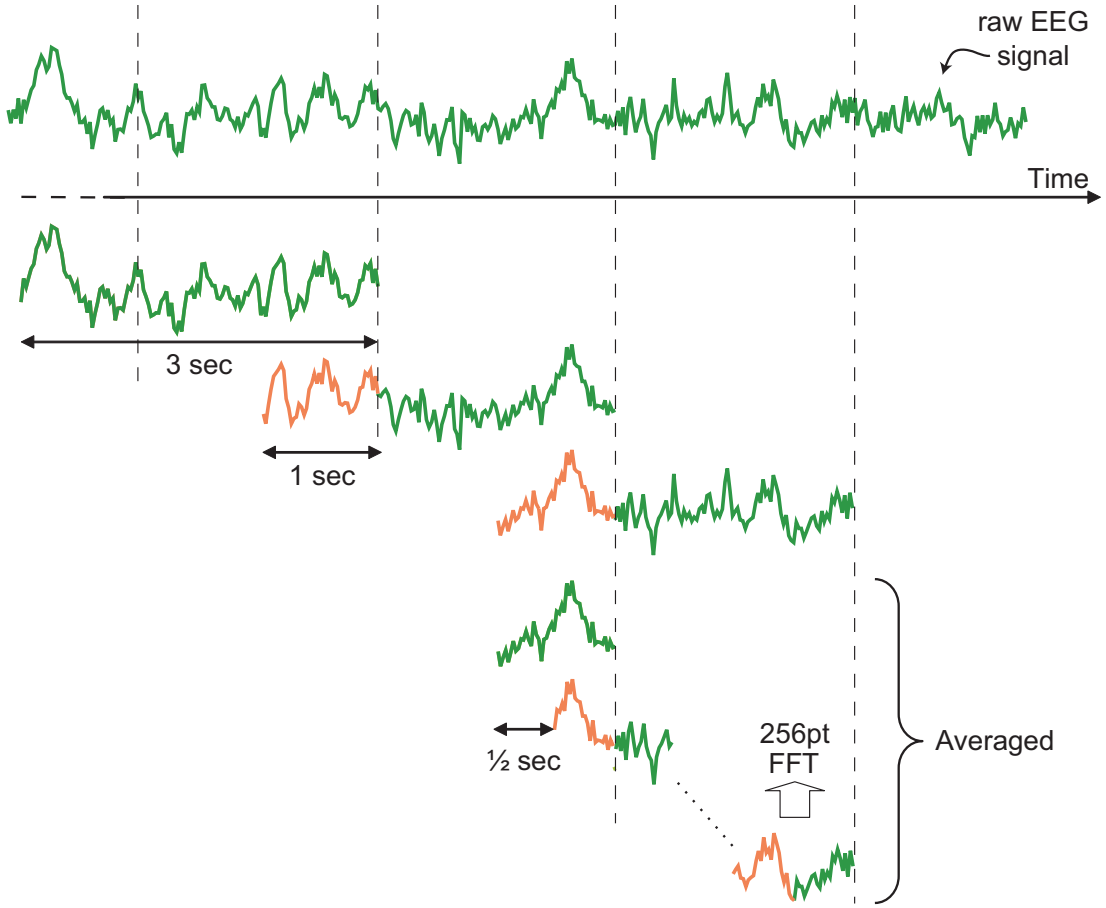


Figure 3.5. Preprocessing EEG data for feature extraction is done by calculating a moving average of overlapping windows of predefined length (3[sec]).

$200[Hz]$ from 12 EEG placements for 2×10 seconds and for each one of the three subjects.

To prepare the collected data for the training of the map, I calculate the moving averaged power spectra of 10×2×7=140 signal sources for the first case and 12×2×3=168 signal sources for the second case. As shown in **Figure 3.5**, I apply a 3[sec](600-point) Hanning window on the signal with a 1[sec](200-point) overlap. The windowed 3[sec] epochs are further subdivided into several 1[sec](200-point) sub-windows using the Hanning window again with 1/2[sec](100-point) overlap, each extended to 256 points by zero padding for a 256-point

fast Fourier transform(FFT). A moving median filter is then applied to average and minimize the presence of artifacts in all the sub-windows. The resulting moving averaged power spectrum is then reduced to six (6) features by integrating the spectral values weighted by six raised-sine shaped windows with an area normalized to unity. Consequently, the feature components do not need to be normalized. In a similar way to the work of Joutsiniemi et al., the weighting windows are overlapping to ensure a smooth change of the features in accordance with the change in power spectrum [31]. These windows cover the following frequency bands:

- Delta : $0 - 4[Hz]$
- Theta : $4 - 8[Hz]$
- Low-Alpha : $8 - 11[Hz]$
- High-Alpha: $11 - 13[Hz]$
- Beta : $13 - 30[Hz]$
- Gamma : $30 - 50[Hz]$

Notice that the alpha band is divided into low and high band. This will allow us to give a different importance coefficient to each frequency band feature, according to its contribution in recognizing the effect of bodily expressions. The resulting time series of EEG power spectrum features consisted of a vector of $10 \times 6 = 60$ features every 2[sec] time intervals for the robot bodily expression case, and a vector of $12 \times 6 = 72$ features every 2[sec] time intervals for the human bodily expression. In each training run, 80% of the resulting feature vectors was randomly chosen. The remaining 20% was used to test the recognition performance.

3.4.3 Training parameters

The 2D map to be learned using the collected features is arranged as a 2D lattice, with each location containing a 60 or 72 dimensional prototype vector. During

Table 3.1. Recognition rates for the case of robot bodily expressions when using data from different sources (subjects)

Number of sources	1	2	3	4	5	6	7
Recognition rate (%)	79.5	74.2	70.8	71.2	70.2	68.6	65.2

the learning process or the self-organization, the importance coefficient w_j, used in the similarity metric (3.4), were taken as 0.5, 0.5, 1.0, 0.9, 0.5, and 0.3 for the features delta, theta, low-alpha, high-alpha, beta, and gamma, respectively. Higher importance was given to the alpha-band feature with an emphasis on the low-alpha band feature, as a result of its sensitivity to the category of bodily expression observed by subjects (see Chapter 2). On the other hand, the higher frequency gamma band feature was given the lowest importance coefficient, since it was not explicitly proven to react to social cues [8].

After training the map, an approximation of the probability density of the input data is reached, generating clusters which can be identified as associated to any of these experimental condition: observing unpleasant bodily expression, observing pleasant bodily expression, or baseline condition.

3.4.4 Recognition when observing Robot bodily expressions

Eighty percent (80%) of the data was used for map training and the remaining 20% was used for evaluation. The resulting recognition rate was of 65.2%, divided into 62.8% for data associated to the observation of unpleasant bodily expressions, 59.3% for data associated to the observation of pleasant bodily expressions, and 73.5% for data associated to the baseline. Clearly, the rate of 65.2% is not satisfactory since this is a low rate to rely on when making a decision. However, when data from a single subject was used in the learning process the recognition jumped to the high rate of 79.5%. The previously obtained low rate of 65.2% is explained by the existence of conflicting data collected from different subjects.

In order to understand the effect of using data from different sources or

Table 3.2. Recognition rates for the case of human bodily expressions when using data from different sources (subjects)

Number of sources	1	2	3	4	5	6	7
Recognition rate (%)	78.9	74.1	71.6	70.8	70.1	69.7	66.5

subjects on the recognition rate, all possible combinations of data sources were used to learn several maps and the recognition rates were calculated. The change in recognition rates is summarized in **Table 3.1**. It can be noticed that the addition of a new source of data decreases the recognition rate by about 5% for the first two additions. However, the rate is almost constant when the number of data sources was 3, 4 and 5. But again the rate decreased by a lower factor when adding more sources to reach the value of 65.2%. This decline is explained by the existence of individual differences in the reaction to bodily expressions and probably even the interpretation. The influence of individual differences is also observed in the previous experiment about impressions (see 2.4.2). To cope with this problem, it is recommended to train several SOMs with a small number of data sources and use them all in the recognition process.

3.4.5 Recognition when observing Human bodily expressions

For this case also, 80% of the data was used for training and the remaining 20% was used for evaluation. The resulting recognition rate was 66.5%, divided into 62.2% for data associated to the observation of unpleasant bodily expressions, 60.8% for data associated to the observation of pleasant bodily expressions, and 76.5% for data associated to the baseline. This rate is better than the 65.2% rate achieved with the data of robot bodily expressions as shown in **Table 3.2**. However, the general figure does not differ much between the two cases suggesting a similar reaction of brain activity.

Using data from different sources showed degradation of the recognition rate, similar to the result of the robot, see **Table 3.2**. The addition of one source

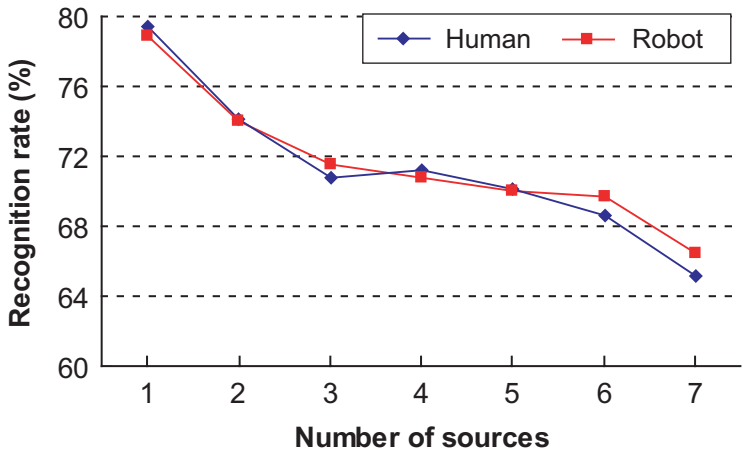


Figure 3.6. Difference in the recognition rates when using different data sources for the robot and the human cases.

resulted in a decrease of 4.8%, then a decrease of 2.5% after adding a third source. The change was somehow stable for the next three additions, but then decreased by 3.2% after the last addition. This result supports the fact that individual differences remain present even though the observed bodily expressions were performed by a human.

3.4.6 Discussion

The recognition rates of the category of the observed bodily expression, regardless of the performer, was about 80% when using data from only one subject. However, this rate decreased significantly when additional data from other subjects was used in the training process. To cope with this problem it would be interesting to train one SOM for each data source, and then combine the resulting SOMs into a bigger structure for the recognition task. Adopting this approach could result in keeping a high recognition rate while taking into consideration all the data that was collected so as not to loose the generality of the solution.

It is interesting to note that the average difference between the recognition rates for robot and human cases is relatively small as shown in **Figure 3.6**. This

proves that SOMs are suitable to the generalization of the effect in the input EEG signals regardless of whether this effect is generated by a human or a robot. Even if differences appear clearly when analyzing raw EEG data, the SOMs succeed in eliminating these differences and keeping only the important information. From this result we can also say that regardless of whether the performer is a robot or a human the brain reaction is similar and so is the recognition rate.

3.5. Recognition with Growing Hierarchical Self-Organizing Maps

3.5.1 Growing Hierarchical Self-Organizing Maps (GHSOM)

The GHSOM[9] enhances the capacity of the basic SOM in two ways. It adopts an incrementally growing version of SOM, which does not require the user to specify the size beforehand, and it adapts to the hierarchical structure in the data by creating a tree-like hierarchy of independent SOMs. The growing capability is meant to represent the data in a homogeneous way on the horizontal plane, while the hierarchy building capability is meant to represent, in more details, data that is cluttered in specific parts of the map. **Figure 3.7** shows an example of a SOM that was trained and labelled using data of *subject1* in experiment 2.4.2. The existence of clusters is emphasized by the gray-scale representation of the distance matrix of the SOM. The labeling sheds some light on the nature of these clusters but it is very complicated to understand the structure of the represented data. It can be noticed also that in some parts of **Figure 3.7** (for instance the upper right part) there is a concentration of big amounts of data, making it almost impossible to grasp the details of the structure at these places. **Figure 3.8** on the other hand, represents a GHSOM trained and labeled using the same data. The resulting GHSOM gives a much better insight into the structure of the data, where specialized clusters are well indicated and deeper levels give more

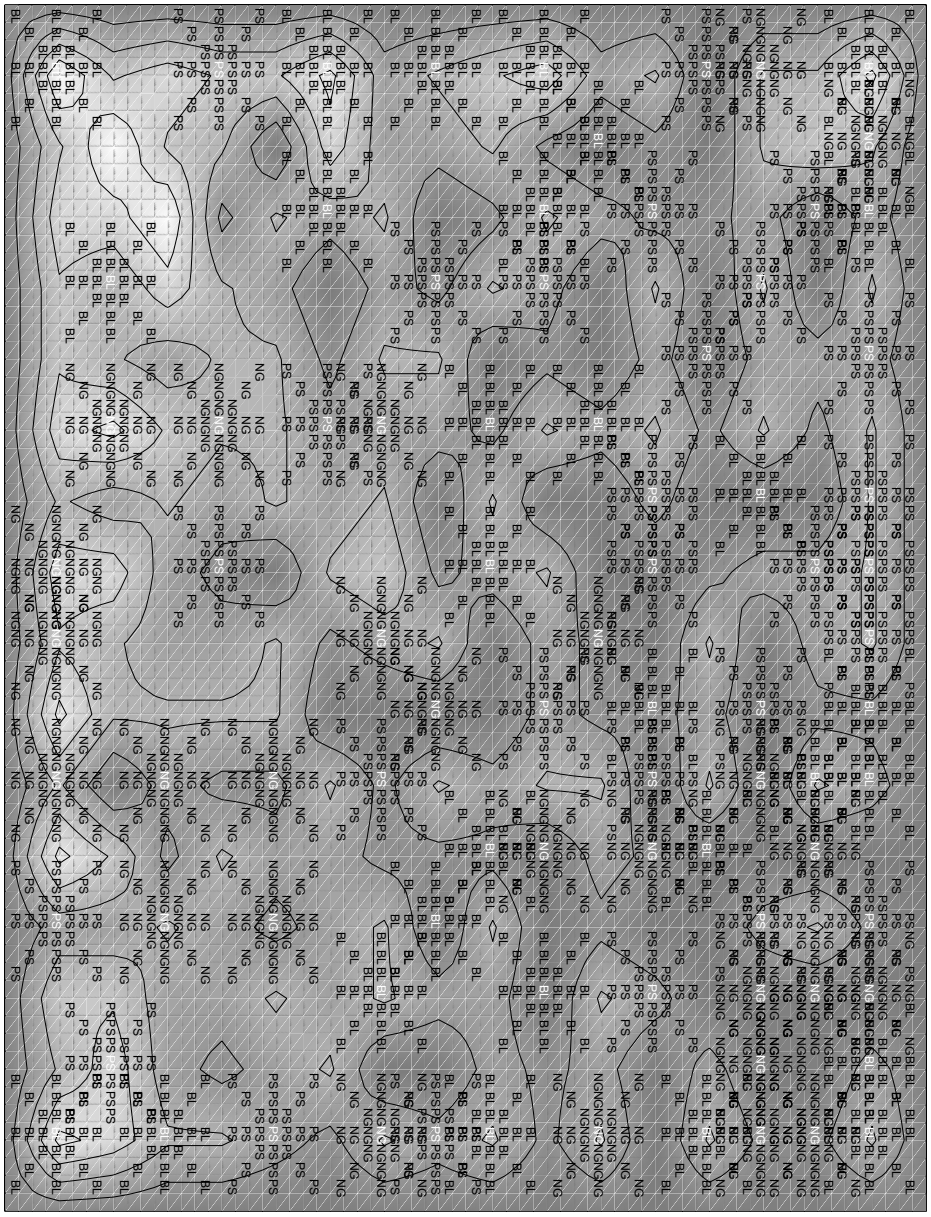


Figure 3.7. SOM trained and labelled using data of subject 1 in experiment 2.4.2, where the distance matrix values are highlighted with gray-scale graduations. The BL label represents baseline units, the NG label represents units associated with unpleasant bodily expressions, and the PL label represents units associated with pleasant bodily expressions.

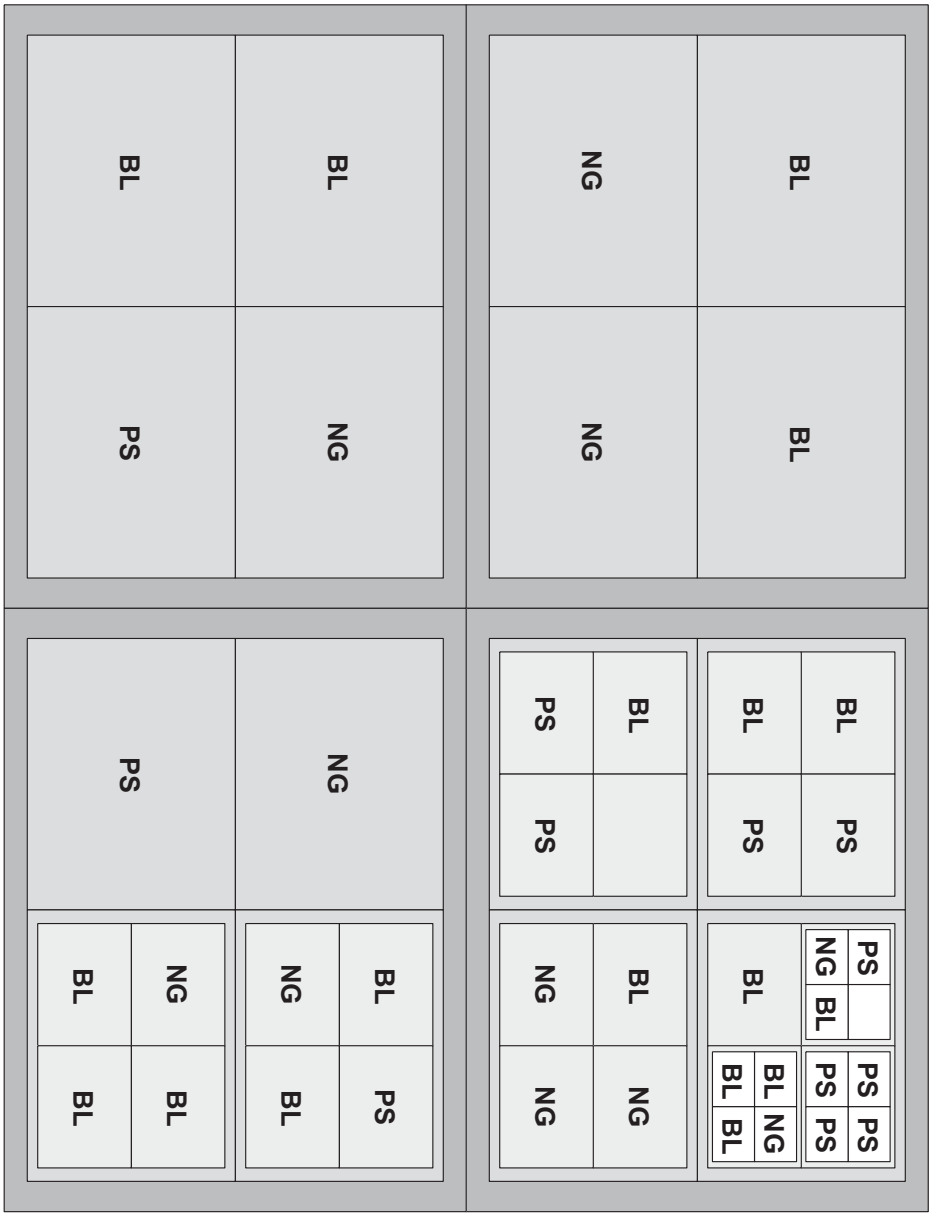


Figure 3.8. GHSOM trained and labeled using data from of subject1 in experiment 2.4.2, where similar gray-scales represents SOMs at the same hierarchical level. The BL label represents baseline clusters, the NG label represents clusters associated with unpleasant bodily expressions, and the PL label represents clusters associated with pleasant bodily expressions.

details to big data clusters, for instance the upper right part.

The basic idea is to create a first layer (root layer) of the hierarchy with one SOM. This SOM is trained with the conventional training algorithm for λ steps. Then, a decision is made if a new row or column needs to be inserted on this level (the growing step). After that, a decision is made if the quality of the map needs to be refined on the next hierarchical level (the expansion step). So, a new 2×2 SOM is added if necessary for every unit on this first layer. The newly created SOMs are trained with the input vectors that were mapped on the parent unit. The prototype vectors of the newly created SOMs are initialized to mirror prototype vectors of neighboring units of the parent unit. When neither the growing step nor the expansion step is necessary then the training process is terminated for the GHSOM.

Size adaptation

The growing step is decided based on the evaluation of whether the current SOM represents the input data properly. The quantization error is used to measure the current representation level or the dissimilarity of all input data mapped onto a particular unit. For each unit i the mean quantization error is calculated by:

$$mqe_i = \frac{1}{|U_i|} \sum_{k \in U_i} D_2 \left(\mathbf{m}_i, \mathbf{x}_k\right), \tag{3.5}$$

where D_2 is the Euclidean distance measure defined by (3.4), U_i is the set of all input vectors mapped to the unit i and $|U_i| \neq 0$. The starting point is to calculate the overall dissimilarity of the input data with the single unit SOM layer 0. This unit is assigned a prototype vector $\mathbf{m}_0$ computed as the average of all the input data using:

$$\mathbf{m}_0 = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \mathbf{x}_i \tag{3.6}$$

where $\mathcal{I}$ is the input data set. The dissimilarity level is then calculated by:

$$mqe_0 = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} D_2 \left(\mathbf{m}_0, \mathbf{x}_i\right) \tag{3.7}$$

The minimum quality of data representation of each SOM will be specified as a fraction of mqe_0 indicated by parameter τ_g. All SOMs must represent their respective data subsets at a mean quantization error smaller than the fraction τ_g of mqe_0, i.e. every unit i should satisfy the following criterion:

$$MQE_m < \tau_g \cdot mqe_0 \tag{3.8}$$

where MQE_m is the mean quantization error for the SOM m, calculated by:

$$MQE_m = \frac{1}{|\mathrm{M}|} \sum_{i \in \mathrm{M}} mqe_i, \tag{3.9}$$

and M is the set of all the units of the map m. If the inequality 3.8 is not fulfilled for the map m then a new row or column is inserted. The unit e with the largest mqe_i is called the error unit, while its most dissimilar unit d has the smallest mqe_i among all the neighbors of e; i.e.:

$$e = \arg\max_i \left(\sum_{j \in C_m} D_2 \left(\mathbf{m}_i, \mathbf{x}_j \right) \right), \quad C_m \neq 0 \tag{3.10}$$

and

$$d = \arg\max_i \left(D_2 \left(\mathbf{m}_e - \mathbf{m}_i \right) \right), \quad \mathbf{m}_i \in \mathrm{N}_e \tag{3.11}$$

where $C_m \subseteq \mathcal{I}$ is a subset of input data associated to the map m, and N_e is a set of the neighboring units of unit e. The new row or column is then inserted between the units e and d. The prototype vectors of the newly inserted units are initialized as the average of their corresponding neighbors.
Alternatively, the quantization error QE defined by:

$$QE_m = \sum_{i \in \mathrm{M}} qe_i, \tag{3.12}$$

may be used to evaluate the quality of data representation of a map m. Where qe_i is defined by:

$$qe_i = \sum_{k \in U_i} D_2 \left(\mathbf{m}_i, \mathbf{x}_k \right) \tag{3.13}$$

In this case, the map growing criterion becomes:

$$QE_m < \tau_g \cdot qe_0 \tag{3.14}$$

Hierarchy adaptation

The expansion step consists of checking for every unit whether the input data mapped to it is inhomogeneous or showing dissimilarities. Therefore, new units are needed to provide more space for better data representation. Sometimes high dissimilarities are not present but rather a big amount of data is concentrated in one unit, resulting in a low level of representation. This step starts by testing the fulfillment of one of the following criterion:

$$mqe_m < \tau_e \cdot mqe_0 \tag{3.15}$$

or

$$qe_m < \tau_e \cdot qe_0 \tag{3.16}$$

depending on which evaluation criteria is adopted for the representativity of input data. Each unit i not fulfilling the criterion will be subject to expansion by creating a new SOM belonging to a sub-layer that will receive as training input the data subset that was mapped to the parent unit i. The prototype vectors of the newly created 2×2 SOM are initialized to mirror the average of the neighboring units of the parent unit. This can be thought of as moving the parent vector in the data space by a fraction toward the selected neighbors; three neighbors for each of the four corner units.

The training is repeated again for all the SOMs on the new layer, followed by an evaluation for growing and expansion until no further action is necessary.

3.5.2 Feature selection

Similar to feature selection in the SOM case (see section 3.4.2) the same approach is adopted for the preprocessing of EEG data in order to prepare enough data points for both the training and the testing steps. For the recognition of the impressions of robot bodily expressions I used the data described in 2.4.2, which was recorded from 10 electrode placements for twice 14 seconds and for each of the seven subjects. As for the recognition of the impressions of human bodily expressions I used the data described in 3.3, which was recorded from 12 electrode

Table 3.3. Recognition rates with GHSOM for the case of robot bodily expressions when using data from different sources (subjects)

Number of sources	1	2	3	4	5	6	7
Recognition rate (%)	83.65	84.95	83.84	83.30	82.68	82.65	82.06

placements for twice 10 seconds and for each of the three subjects. The resulting time series of EEG power spectrum features consisted of a vector of 60 features every 2[sec] time interval for the robot bodily expressions case, and a vector of 72 features every 2[sec] time interval for the human bodily expressions case. In each training run, 80% of the resulting feature vectors were randomly chosen; while the remaining 20% were used to test the recognition performance.

3.5.3 Training parameters

The dimensions of the prototype vectors associated to each unit in the learned GHSOM were 60 and 72 for the robot and the human case respectively. During the learning process, the importance coefficient w_j, used in the similarity measure (3.4), were kept similar to the ones used for SOM training (see 3.4.3). Their values were 0.5, 0.5, 1.0, 0.9, 0.5, and 0.3 for the features delta, theta, low-alpha, high-alpha, beta, and gamma, respectively. Giving higher importance to the alpha frequency bands due to their proven reaction to bodily expressions; while other frequency bands were given lower importance coefficients.

Once the training is finished, clusters appeared on the resulting GHSOMs. These ones were labeled using a small amount of data previously labeled by hand. The resulting clusters were associated with one of the experimental conditions: observing pleasant bodily expression, observing unpleasant bodily expressions, or baseline condition.

Table 3.4. Recognition rates with GHSOM for the case of human bodily expressions when using data from different sources (subjects)

Number of sources	1	2	3	4	5	6	7
Recognition rate (%)	83.20	83.95	83.90	83.76	83.43	83.19	82.87

3.5.4 Recognition when observing Robot bodily expressions

Eighty percent (80%) of the data was used for training and the remaining 20% was used for evaluation. The resulting recognition rate was of 82.06%, divided into 85.24% for data associated to the observation of pleasant bodily expressions, 64.29% for data associated to the observation of unpleasant bodily expressions, and 96.65% for data associated to the baseline. The rate of 82.06% is satisfactory since the data used for training and recognition was not cleaned of artifacts. This is very important because in order to adapt the robot's behavior there is a need to identify correctly the mental activity and to take into consideration the existence of noisy data.

In order to understand the effect of using data from different sources or subjects on the recognition rate, all possible combinations of data source were used to learn several GHSOM maps and the recognition rates were calculated. The change in recognition rates is summarized in **Table 3.3**. It can be noticed that the gradual addition of new sources decreases the recognition rate by a very small amount. There is a small change in the recognition rate with an average variation of about 0.37%. The individual differences in the reaction to bodily expressions are believed to be the cause of this variation in the recognition rate, but their effect is very small.

3.5.5 Recognition when observing Human bodily expressions

In this case also 80% of the data was used for training and the remaining 20% was used for evaluation. The resulting recognition rate was of 82.87%, divided into 81.4% for data associated to the observation of pleasant bodily expressions, 78.2% for data associated to the observation of unpleasant bodily expressions, and 89.0% for data associated to the baseline. This is slightly better than the 82.06% rate achieved with the data of robot bodily expressions as shown in **Table 3.3**. However, the general figure is still similar between the two cases suggesting the similarity in brain activity reaction.

Using data from different sources showed no big change in the recognition rate, similar to the result for the robot case (see **Table 3.4**). The addition of one source resulted in an increase of the recognition rate by 0.75%, then a decrease by 0.05% resulted after adding a third source, and this goes on in almost a constant value. The first increase is explained as the effect of having more data for the training. This results in a better representation of the input data and thus a better recognition rate. It can be concluded that the amount of data used when considering only one source was not enough to represent properly the underlying structure in the input data. However, after adding a third source the recognition rate decreased as expected due to individual differences. Nevertheless, this rate was still higher than the rate of one data source, confirming again the under representativity of the amount of data extracted from only one source.

3.5.6 Discussion

It is interesting to notice that the average difference between the recognition rates of robot and human cases is relatively small as shown in **Figure 3.9**. It is even smaller than the same result obtained with SOM. This proves that GHSOMs are suitable for the generalization of the effect in the input EEG signals regardless of whether this effect is generated by the observation of a human or a robot performer. Even if differences appear clearly when analyzing raw EEG data,

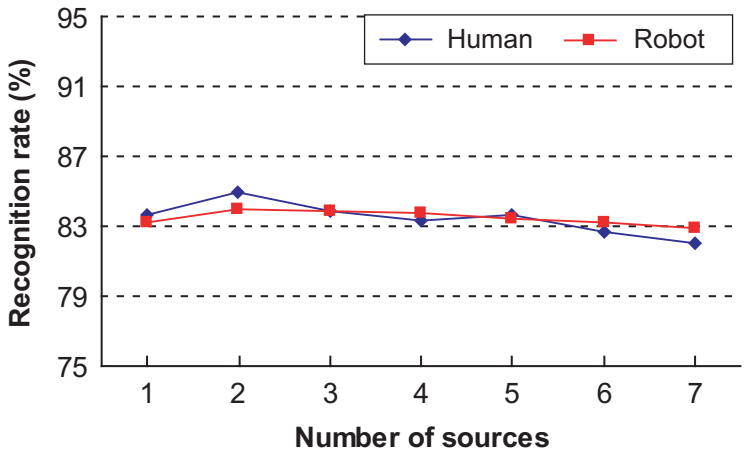


Figure 3.9. Difference in the recognition rates with GHSOM when using different data sources for the robot and the human cases.

the GHSOMs succeeded in eliminating these differences and keeping only the important information.

The recognition rates of the category of the observed bodily expression, without considering the performer, were of about 83% when using data from only one subject. This rate remained relatively stable when additional data from other subjects was used in the training process. Compared to the results obtained with the conventional SOM, the results obtained with GHSOM are more robust to individual differences (see **Figure 3.10**). This is a characteristic of prime importance that could lead to a user-free recognition of brain reactions to similar stimuli. The hierarchization of several SOMs into a bigger structure allowed the handling of inherent differences in the input data due to individual differences, while keeping the recognition rate at its highest level.

3.6. Conclusion

In this chapter, I presented two computational methods to use for the recognition of the impressions of bodily expressions. Both the self-organizing maps (SOM) and the growing hierarchical self-organizing maps (GHSOM) were used for the

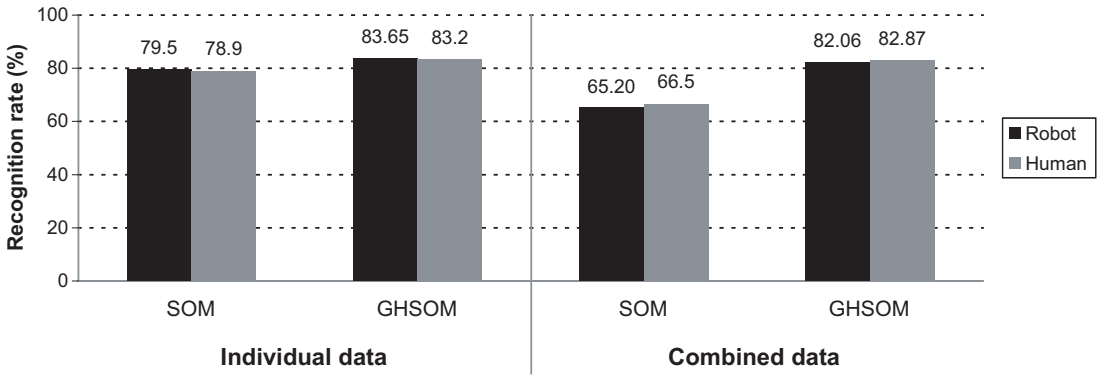


Figure 3.10. Difference in the recognition rates with SOM and GHSOM when using different data sources for the robot case.

recognition. It was shown that SOM achieved relatively high recognition rates considering that the data which was used was not filtered for noise elimination. However, the GHSOM gave better results and improved the recognition rate by almost 5%, while giving an insight about the underlying structure of the data. Its growing characteristic gave it the ability to spread and smooth data on one layer of the hierarchy; whereas, its expansion characteristic gave it the ability to represent details about the complicated parts of the data. The result is that GHSOM can cope perfectly with the individual differences present in the reaction of the data collected from different subjects, while maintaining an almost constant recognition rate.

To complete the results of the previous chapter about brain reaction to robot bodily expressions, an additional experiment was conducted to collect data when subjects observed human bodily expressions. This data was also used for the recognition task and gave similar results to the data collected for robot bodily expressions. This confirms that an observer gets impressed in a similar way when observing bodily expressions regardless of the whether the performer is a human or a robot. Although this result is different from previous research results which state that the studied parts of the human brain react only to biological motions, it suggests that robots are close enough to humans such that they generate the same effect. This brings up an interesting interrogation about the characteristics

in a performer that activate the social perception mechanism in humans.

Future research directions should focus on improving the recognition rate to a higher degree and to try to recognize a refined classification of bodily expressions. A link with human motion styles [24] would be interesting to provide more details about the bodily expressions. Another direction would be to link the bodily expressions to emotions or affect[46]. There is an extensive work on generating emotional motions [3, 39] that could be incorporated in this research.

Chapter 4

Conclusion and future work

The goal of this research work is to investigate the relation between bodily expressions executed by a robot and their impression on an observer. Understanding this relation and being able to assess its state would allow the creation of an adaptive behavior for robots; and thus maximize their probability of being accepted within society. The most direct approach to study this problem is to use reliable physiological measurements that react to the observation of bodily expressions. The brain happens to be the organ involved in the interpretation of observed events. Thus analyzing measured brain activity was necessary for this study. I adopted the relatively new but well established framework of brain-machine interface (BMI) to tackle this problem. This involved answering the following key questions: What kind of visual stimuli to generate? What measurement technique to use? Which features to select? Where to record from in the brain? and finally, How to infer the internal state of the brain?

In chapter two, I investigated the relation between bodily expressions and their impressions on an observer. I started by generating six bodily expressions, then I classified them into to two categories according to their expressiveness (pleasant and unpleasant). Their expressiveness was confirmed statistically with a self-reporting experiment where a number of volunteers answered questionnaires about the bodily expressions. Afterwards, I conducted an experiment to assess the impressions on the observers while watching the considered bodily expressions

by collecting the observer's brain activity using electroencephalogram (EEG). Previous research findings in the field of cognitive science were used to determine the brain regions that needed to be monitored. The superior temporal sulcus (STS), which reacts to social cues, and the mirror neurons located in the prefrontal cortex (PFC) that get activated during imitation and learning tasks, were the regions of interest in this study. The method adopted for spectral analysis revealed a correlation between the power level of low-alpha (8-11$[Hz]$) frequency band and the category of the observed bodily expression. The reproducibility or repeatability of this band's reaction was confirmed with a third experiment where a subject observed candidate bodily expressions for each category. Also, the reaction of the power level of low-alpha frequency band proved to be inversely proportional to the category of the observed bodily expression.

In chapter three, I presented two computational methods to use for the recognition of the impression of bodily expressions. Both the self-organizing maps (SOM) and the growing hierarchical self-organizing maps (GHSOM) were used for the recognition. It was shown that SOM achieves relatively high recognition rates considering that the data used is not filtered for noise elimination. However, the GHSOM gives better results and improves the recognition rate by almost 5%, while giving an insight about the underlying structure of the data. Its growing characteristic gives it the ability to spread and smooth data on one layer of the hierarchy; whereas, its expansion characteristic gives it the ability to represent details about the complicated parts of the data. The major result is that GHSOM can cope perfectly with the individual differences present in the reaction of the data collected from different subjects, while maintaining an almost constant recognition rate. Also, the non-artifact-free data, which were used in the training as well as the recognition, showed the effectiveness of this method when planning to close the BMI loop and give a biofeedback to the observer according to his/her estimated state of mind.

To complete the results of chapter two about brain reaction to robot bodily expressions, an additional experiment was conducted to collect data when subjects observed human bodily expressions. This data was also used for the

recognition task and gave similar results to the data collected for robot bodily expressions. This confirms that an observer gets impressed in a similar way when observing bodily expressions regardless of the whether the performer is a human or a robot.

Future research directions should focus on the degree to which brain reaction appears when observing robots with different human-like physical and behavioral characteristics. The reaction of the power level of low alpha frequency band has opened the opportunity to utilize this feature to examine the capacity of a humanoid robot in activating the social perception system in a human observer. Furthermore, the understanding of which robot properties are necessary or sufficient to activate the social perception system in an observer is of particular interest, since regardless of the performer of the chosen bodily expressions, the robot or the human, there was no significant difference between the recognition rates for both cases.

Moreover, there is a need to focus on improving the recognition to a bigger rate and to try to recognize a refined classification of bodily expressions. Depending on the requirements of the considered application the number of the bodily expression categories will vary and thus influence the overall recognition rate. A link with human motion styles [24] would be interesting to provide more details about the bodily expressions. Another direction would be to link the bodily expressions to affect[46]. There is an extensive work on generating emotional motions [3, 39] that could also extend this research work.

List of Publications

Major Publications

Journal Articles *(peer reviewed)*

- **A. Khiat**, M. Toyota, Y. Matsumoto, and T. Ogasawara, *"Assessment of the Impressions of Robot Bodily Expressions using Electroencephalogram Measurement of Brain Activity"*, a chapter in "Humanoid Robots" book, International Journal of Advanced Robotic Systems, April 2007. *(Accepted)*

International Conferences *(peer reviewed)*

- **A. Khiat**, M. Toyota, Y. Matsumoto, and T. Ogasawara. *"Brain Activity in the Evaluation of the Impression of Robot Bodily Expression"*, in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006), Beijing (China), pp.5504-5508, Oct. 9-15, 2006.

- **A. Khiat**, M. Toyota, Y. Matsumoto, and T. Ogasawara. *"Investigating the Relation between Robot Bodily Expressions and their Impression on the User"*, in Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI2006), Sydney (Australia), pp.339-341, Jan. 29-Feb. 1, 2006.

National Conferences *(without review)*

- **A. Khiat**, *"Assessment of the Impression of Robot Bodily Expressions using Electroencephalogram Measurement of Brain Activity"*, NAIST-IS COE Symposium, Tokyo, Dec. 8, 2006.

- **A. Khiat**, *"Brain Activity in the Evaluation of the Impression of Robot Bodily Expressions"*, NAIST-IS COE Festival, Nara, Mar. 23, 2006.

- 豊田 将隆, **キアット・アブデラズィズ**, 小笠原 司, 松本 吉央,"ロボットインタラクションのための脳波を用いた人間の感情推定", ロボティクスメカトロニクス講演会 2005 (ROBOMEC2005), 2P1-N-029, 6 月 2005 年.

Other Publications

Journal Articles *(peer reviewed)*

- R. Ouiguini, **A. Khiat** and S. Ouzaid, *"Local Maneuver Planning: Application to an Autonomous Wheelchair Moving in Constrained Spaces"*, in Advanced Robotics: the RSJ International Journal, Special Issue on Rehabilitation Robots, vol. 14, no. 7, pp.607-623, 2000.

International Conferences *(peer reviewed)*

- **A. Khiat**, S. Yous, M. Kidode, and T. Ogasawara, *"Combining Fixed Stereo and Active Monocular Cameras in a Platform for Security Applications"*, in Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO2006), pp.1134-1139, Kunming (China), Dec. 17-20, 2006.

- S. Yous, **A. Khiat**, M. Kidode, and T. Ogasawara, *"Networked Heterogeneous Camera System for High Resolution Face Images"*, in Proceedings of the 2nd International Symposium on Visual Computing (ISVC2006), LNCS 4292, pp.88-97, Lake Tahoe (USA), Nov. 6-8, 2006.

- **A. Khiat**, Y. Matsumoto, and T. Ogasawara, *"Task Specific Eye Movements Understanding for a Gaze-Sensitive Dictionary"*, in Proceedings of The 8th ACM International Conference on Intelligent User Interfaces (IUI2004), Island of Madeira (Portugal), pp.265-267, Jan. 13-16, 2004.

- **A. Khiat**, Y. Matsumoto, and T. Ogasawara, *"Vers une Assistance Proactive dans l'Interaction Homme-Machine basée sur l'Analyse du Point de Regard"*, in CD-ROM Proceedings of Journées Scientifiques Francophones (JSF2003), Tokyo (Japan), Nov. 24-26, 2003.

- **A. Khiat**, Y. Matsumoto, and T. Ogasawara, *"Toward Gaze-based Proactive Support for Web Readers"*, in Proceedings of the 12th IEEE International Workshop on Robot and Human Communication (RO-MAN2003), California (USA), pp.97-103, Oct. 31-Nov. 2, 2003.

- Y. Adachi, K. Goto, **A. Khiat**, Y. Matsumoto, and T. Ogasawara, *"Estimation of User's Attention based on Gaze and Environment Measurements for Robotics Wheelchair"*, in Proceedings of the 12th IEEE International Workshop on Robot and Human Communication (RO-MAN2003), Silicon-Valley (USA), pp.97-103, Oct. 31-Nov. 2, 2003.

- H. Takemura, **A. Khiat**, H. Iwama, J. Ueda, Y. Matsumoto, and T. Ogasawara, *"Study of the Toes Role in Human Walk by a Toe Elimination and Pressure Measurement System"*, in Proceedings of The IEEE International Conference on Systems, Man & Cybernetics (SMC2003), Washington D.C. (USA), pp.2569-2574, Oct. 5-8, 2003.

National Conferences *(without review)*

- S. Yous, **A. Khiat**, and H. Laga, *"Support System for Visual Monitoring"*, NAIST-IS COE Symposium, Tokyo, Dec. 8, 2006.

- **A. Khiat**, and S. Yous, *"A Networked Platform of Fix Stereo Cameras and Active Monocular Cameras for Security Applications"*, Pervasive Comput-

ing Symposium, Nara, Jul. 5, 2006.

- **A. Khiat**, and S. Yous, *"Networked Camera System for Face Tracking"*, NAIST-IS COE Festival, Nara, Mar. 23, 2006.

- 麻生　英樹, 柿倉　正義, 小玉　智志, **キアット・アブデラジズ**, 松本　泰明, 本村　陽一, 原　功, 浅野　太, 小笠原　司, 新田　恒雄,"確率的推論を利用したマルチモーダル対話制御", 第 17 回 人工知能学会全国大会 2003 (JSAI 2003), 1C1-04, 6 月 2003 年.

Acknowledgements

I would like to express my gratitude and special thanks to Professor Tsukasa Ogasawara for all his support and guidance over the past six years and throughout my stay in his laboratory at NAIST.

My deepest gratitude goes to Professor Kenichi Matsumoto of NAIST for sparing me his precious time to evaluate my work with interest, and for letting me utilize the specialized equipment in his laboratory as a part of this thesis work.

Many thanks go to Assistant Professor Etsuko Ueda of NAIST for her support and fruitful discussions about the subject and the methods adopted to solve the considered problem.

Particular thanks go to Assistant Professor Jun Ueda of NAIST (currently at MIT) for pushing me toward Brain-Machine Interfaces and for introducing me to the members of Osaka University dealing with this subject.

Much appreciation goes to Assistant Professor Masayuki Hirata and Associate Professor Amami Kato of the Medical School of Osaka University, for the very interesting discussion about BMI and their pioneering work with EelectroCorticography analysis of human motion.

Particular thanks go to Masataka Toyota (currently at Cannon Corporation)

for starting this project with me and for spending a whole year of struggle and active endeavor.

Special thanks go to Masahiro Kondo and Sofiane Yous for their friendship and their support during the particularly depressing moments. Without their help many parts of this work would not have been accomplished properly.

Distinguished thanks go to Albert Causo for being always ready to discuss my work and to check my writings. Without his help I would have had much trouble expressing my thoughts and ideas correctly,

All my heartfelt appreciation goes to the current and previous students of Robotics laboratory at NAIST for their help and kindness. Without them I would not have enjoyed thoroughly my stay.

Special thanks to all my friends for the moral support, their patience during my difficult moments and their cheerfulness during joyful ones.

Particular thanks go to my parents and family, whose unwavering love and support goes beyond limits and provides me with strength to pursue my studies.

Finally, I would like to thank the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the NAIST-IS COE "Ubiquitous and Media Computing" program for supporting me financially during all my period of stay in Japan, giving me a chance to discover new, different and interesting culture.

References

[1] Truett Allison, Aina Puce, and Gregory McCarthy. Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, 4(7):251–291, 2000.

[2] Shun-Ichi Amari. Topographic organization of nerver fields. *Bulletin of Mathematical Biology*, 42(3):339–364, 1980.

[3] Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from motion. In *Graphics Interface '96*, pages 222–229, 1996.

[4] Michael R. Anderberg. *Cluster Analysis for Applications (Probability & Mathematical Statistics Monograph)*. Academic Press, USA, 1973.

[5] Irmgard Bartenieff and Dori Lewis. *Body Movement: Coping with the Environment*. Gordon and Breach Science Publishers, USA, 1980.

[6] Eva Bonda, Michael Petrides, David Ostry, and Alan Evans. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, 16(11):3737–3744, 1996.

[7] Cynthia Breazeal. *Designing Sociable Robots*. The MIT Press, USA, 2002.

[8] S. Cochin, C. Barthlemy, B. Lejeune, S. Roux, and J. Martineau. Perception of motion and qEEG activity in human adults. *Electroencephalography and Clinical Neurophysiology*, 107:287–295, 1998.

[9] Michael Dittenbach, Andreas Rauber, and Dieter Merkl. Uncovering hierarchical structure in data using the growing hierarchical self-organizing map. *Neurocomputing*, 48:199–216, 2002.

[10] Jerome H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Society*, 82:249–266, 1987.

[11] Jerome H. Friedman and John W.Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23(9):881–890, 1974.

[12] Allen Gersho. Asymptotically optimal block quantization. *IEEE Transactions on Information Theory*, 25(4):373–380, 1979.

[13] S.T. Grafton, M.A. Arbib, L. Fadiga, and G. Rizzolatti. Localization of grasp representation in humans by positron emission tomography: 2. observation compared with imagination. *Experimental Brain Research*, 112:103–111, 1996.

[14] Robert M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, 1984.

[15] Julie Grezes, Nicolas Costes, and Jean Decety. Top-down effect of strategy on the perception of human biological motion: a PET investigation. *Cognitive Neuropsychology*, 15(6/7/8):553–582, 1998.

[16] Julie Grezes, Nicolas Costes, and Jean Decety. The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain*, 122(10):1875–1887, 1999.

[17] E. Grossman, M. Donnely, R. Price, D. Pickens, V. Morgan, G. Neighbor, and R. Blake. Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12(5):711–720, 2000.

[18] John A. Hartigan. *Clustering Algorihtms*. John Wiley & Sons, USA, 1973.

[19] Trevor Hastie and Werner Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.

[20] Leigh R. Hochberg, Mijail D. Serruya, Gerhard M. Friehs, Jon A. Mukand, Maryam Saleh, Abraham H. Caplan, Almut Branner, David Chen11, Richard D. Penn, and John P. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442:164–171, 2006.

[21] Elizabeth A. Hoffman and James V. Haxby. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3:80–84, 2000.

[22] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[23] R.J. Howard, M. Brammer, I. Wright, P.W. Woodruff E.T. Bullmore ET, and S. Zeki. A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Current Biology*, 6(8):1015–1019, 1996.

[24] Eugene Hsu, Kari Pulli, and Jovan Popović. Style translation for human motion. *ACM Transactions on Graphics*, 24(3):1082–1089, 2005.

[25] Jun'ichi Ido, Kentaro Takemura, Yoshio Matsumoto, and Tsukasa Ogasawara. Robotic receptionist ASKA: a research platform for human-robot interaction. In *IEEE Workshop of Robot and Human Interactive Communication*, pages 306–311, October 2002.

[26] Masato Ito and Jun Tani. On-line imitative interaction with a humanoid robot using a dynamic neutral network model of a mirror system. *Adaptive Behavior*, 12(2):93–115, 2004.

[27] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, USA, 1988.

[28] Nicholas Jardine and Robin Sibson. *Mathematical Taxonomy (Probability & Mathematical Statistics S.)*. John Wiley & Sons, USA, 1971.

[29] Herbert Henry Jasper. The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375, 1958.

[30] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973.

[31] Sirkka-Liisa Joutsiniemi, Samuel Kaski, and T. Andreo Larsen. Self-organizing map in recognition of topographic patterns of EEG spectra. *IEEE Transactions on Biomedical Engineering*, 42(11):1062–1068, 1995.

[32] John W. Sammon Jr. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, 1969.

[33] Samuel Kaski. *Data Exploration using Self-Organizing Maps*. PhD thesis, Helsinki University of Technology, 1997.

[34] Abdelaziz Khiat, Masataka Toyota, Yoshio Matsumoto, and Tsukasa Ogasawara. Brain activity in the evaluation of the impression of robot bodily expressions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5504–5508, 2006.

[35] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernitics*, 43:59–69, 1982.

[36] Z. Kourtzi and N. Kanwisher. Activation in human mt/mst by static images with implied motion. *Journal of Cognitive Neuroscience*, 12(1):48–55, 2000.

[37] Hideki Kozima. *Infanoid: A babybot that explores the social environment*, pages 157–164. Socially Intelligent Agents: Creating Relationships with Computers and Robots. Kluwer Academic Publishers, The Netherlands, 2002.

[38] Eric C. Leuthardt, Gerwin Schalk, Jonathan R. Wolpaw, Jeffrey G. Ojemann, and Daniel W. Moran. A brain-computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, 1(2):63–71, 2004.

[39] Hunok Lim, Akinori Ishii, and Atsuo Takanishi. Emotion-base biped walking. *Robotica*, 22:577–586, 2004.

[40] James B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1966.

[41] John Makhoul, Salim Roucos, and Herbert Gish. Vector quantization in speech coding. *Proceedings of the IEEE*, 73(11):1551–1588, 1985.

[42] Miguel L. Nicolelis. Actions from thoughts. *Nature*, 409(18):403–407, 2001.

[43] Tatsuya Nomura, Takayuki Kanda, and Tomohiro Suzuki. Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI & Society*, 20(2):138–150, 2006.

[44] Donald A. Norman, Andrew Ortony, and Daniel M. Russell. Affect and machine design: Lessons for the development of autonomous machines. *IBM Systems Journal*, 42(1):38–44, 2003.

[45] Jaime Owner A. Pineda and Lindsay M. Oberman. What goads cigarette smokers to smoke? neural adaptation and the mirror neuron system. *Brain Research*, 1121(1):128–135, 2006.

[46] Frank E. Pollick, Helena M. Paterson, Armin Bruderlin, and Anthony J. Sanford. Perceiving affect from arm movement. *Cognition*, 82:B51–B61, 2001.

[47] Aina Puce and Truett Allison. Differential processing of mobile and static faces by temporal cortex. *NeuroImage*, 9:S801, 1999.

[48] Aina Puce, Truett Allison, Shlomo Bentin, John C. Gore, and Gregory McCarthy. Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, 18(6):2188–2199, 1998.

[49] Giacomo Rizolatti and Laila Craighero. The mirror-neuron system. *Annual Reviews of Neuroscience*, 27:169–192, 2004.

[50] G. Rizzolati, L. Fadiga, M. Matelli, V. Bettinardi, E. Paulesu, D. Perani, and F. Fazio. Localization of grasp representations in humans by PET: 1. observation versus execution. *Experimental Brain Research*, 111:246–252, 1996.

[51] Alois Schlögl, Stephen J. Roberts, and Gert Pfurtscheller. A criterion for adaptive autoregressive models. In *IEEE International Conference of Engineering in Medicine & Biology Society*, pages 1581–1582, 2000.

[52] Stefan Shaal. Is imitation learning the route to humanoid robot? *Trends in Cognitive Science*, 3(6):233–242, 1999.

[53] Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O'Connor. Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52(6):1061–1086, 1987.

[54] Kazuhiko Shinozawa, Futoshi Naya, and Kiyoshi Kogure. Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human Computer Study*, 62(2):267–279, 2005.

[55] P.H.A. Sneath and R.R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* W. H. Freeman, USA, 1973.

[56] Kojima Takashi, Tsunashima Hitoshi1, Shiozawa Tomoki, Takada Hiroki, and Sakai Takuji. Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS). *Optical and Quantum Electronics*, 37:1319–1338, 2005.

[57] Toru Tanaka, Taketoshi Mori, and Tomomasa Sato. Quantitative analysis of impression of robot bodily expression based on laban movement theory. *Journal of Robotics Society of Japan*, 19(2):252–259, 2001. (in Japanese).

[58] Kurt A. Thoroughman and Reza Shadmehr. Learning of action through adaptive combination of motor primitives. *Nature*, 407:742–747, 2000.

[59] S.H.C. Du Toit, A.G.W. Steyn, and R.H. Stumpf. *Graphical Exploratory Data Analysis*. Springer-Verlag, USA, 1986.

[60] W.S. Torgerson. Multidimensional scaling: 1. theory and method. *Psychometrika*, 17:401–419, 1952.

[61] Robert C. Tryon and Daniel E. Bailey. *Cluster Analysis*. McGraw-Hill, USA, 1973.

[62] John Wilder Tukey. *Exploratory Data Analysis*. Addison-Wesley, USA, 1977.

[63] Johan Wessberg, Christopher R. Stambaugh, Jerald D. Kralik, Pamela D. Beck, Mark Laubach, John K. Chapin, Jung Kim, S.James Biggs, Mandayam A. Srinivasan, and Miguel L. Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408:361–365, 2000.

[64] Cynthia M. Whissel. *The Dictionary of Affect in Language*, volume 4 of *Emotion: Theory, research and experience. The measurement of emotions*. Academic Press, USA, 1989.

[65] Bruno Wicker, Francois Michel, Marie-Anne Henaff, and Jean Decety. Brain regions involved in the perception of gaze: a PET study. *NeuroImage*, 8(2):221–227, 1998.

[66] Jonathan Wolpaw and Dennis J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences*, 101(51):17849–17854, 2004.

[67] Gale Young and Alston S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.