UNIVERSITÀ DEGLI STUDI DI TRENTO
CIMeC - Center for Mind/Brain Sciences

MASTER DEGREE IN COGNITIVE SCIENCE

Language and Multimodal Interaction Track

THESIS

LAUGHING WITH A VIRTUAL AGENT:
an evaluation study using a virtual character expressing different copying behaviours.

Supervisor:
Prof. Massimo Poesio

Second supervisor:
Prof. Catherine Pelachaud

Student:
Beatrice Biancardi

ACADEMIC YEAR 2013 - 2014

Acknowledgments

A particular thank you to Professor Massimo Poesio, who introduced me to Professor Catherine Pelachaud, and a thank you to her, who welcomed me in her lab and guided me during all my internship, helping me with advices also in the management of the Thesis.

A thank you to Greta team, in particular to Florian, with whom I collaborated during all this months, who explained me many practical things and was always available for exchanging ideas and doubts. A thank you to Angelo, who allowed me not to forget Italy, and supported me in the hard world of Statistics.

I would like to thank my boyfriend Nadir, who supported and tolerated me during the last intense period, always available to help me; a thank you to my family who allowed me to reach this goal; a thank you to my University colleagues, with whom I have shared joys and labour.

Lastly, I would like to thank all people who helped with advices and criticisms, to the realization of this Thesis.

Beatrice Biancardi

Thesis rationale

The main idea which led me while deciding the subject of my Master Thesis was to make a comprehensive work that could take into account and link together the different courses attended during the Master. My final choice thus concerned the theme of Laughter, since it could be analysed following the interdisciplinary approach typical of Cognitive Science. Indeed, it can be studied by very different but sometimes complementary disciplines: you can think about laughter as a form of social communication, you can analyse its phonetic properties and comparing them among different cultures; you can decide to consider laughter within the set of emotions, and concentrate on its psychological aspects, as well as to investigate it more deeply using neuroscientific methods.

The overview proposed in this Thesis will consider the arguments mentioned above, up to concentrate on the most relevant topic, that reflects the characteristic track of my Master, i.e. the Human-Computer interaction. The state of the art in this field regarding Laughter will be introduced, followed by the detailed explanation of the work carried out during my internship experience at Télécom Paris-Tech in Paris, within the group called "Greta team". This team, led by the research director Catherine Pelachaud, has as members PhD students and post-doc people having different backgrounds and complementary skills; they work on making human-computer interactions more emotion-oriented and social, in order to improve communicative skills of virtual agents and learning about human behaviour and communication.

The second reason why I chose Laughter as the subject of my Thesis is the relative lack of materials in scientific research about this topic, with respect to other cognitive processes and emotions; the first interest about laughter came from other disciplines like philosophy, rather than physics or science. However, during the last years, laughter has started to leave its secondary role and to become an argument of interest, also in psychology and in AI. The main research in this field was performed within the European project Ilhaire, started in 2011 and just terminated, which involved a team of researchers from different Universities, all with the aim of improving the theoretical and technical knowledge about laughter, with the final purpose of incorporating it in ECAs. The work described in the last part of the Thesis tries to add a contribute to these researches, inasmuch it concerns endowing a laughing virtual agent with copying behaviours and evaluating its performance during an interaction with the user.

The structure of the Thesis is divided in two parts: the first one aims to introduce laughter and virtual agents, while the second part describes the creation of a laughing virtual agent able to copy user's laughter intensity, and the evaluation of such system.

- Chapter one: it defines human laughter, and its relation with humour. The the main hypotheses about the possible functions of laughter, and its different causes, as well as a brief discussion about the methods used to study humour and laughter. Finally, behavioural patterns and the neural basis of laughter are briefly described.
- Chapter two: it introduces the world of virtual agents, and the more specific Embodied Conversational Agents. After giving definitions and descriptions of them, it analyses the state of the art of applications, and why it is useful to provide ECAs with expressivity and laughter.
- Chapter three: it introduces the ECA Greta, built by Greta team of Télécom Paris-Tech, and the general architecture of the system used for the experiment.
- Chapter four: it describes into details the evaluation study performed during my internship at Télécom Paris-Tech, from the research questions to the final results, passing through the set up, the experimental design and all the other steps of the experiment.

Contents

Part I

Laughter and Virtual Agents: theory and state of the art.

Chapter 1

Laughter in humans

In this chapter, the concept of laughter is presented, under several points of views: its definition and classifications are given in 1.1, an overview of laughter theories is given in 1.2, the methods used to study laughter are summarized in 1.3, the behavioural patterns of laughter are described in 1.4, and finally the neural correlates are illustrated in 1.5.

1.1 Attempts to define and classify laughter

Everybody has experienced, at least once (and I hope more) during life, laughter.

The behavioural pattern of laughter is universal, and recognizable by people from different countries and cultures. In neonates, it emerges later than smiling, around the fourth month, but various evidences support the hypothesis of its innateness, such as the presence of all the functional structures related to laughter since the date of birth, confirmed by cases of gelastic epilepsy, and the observation of laughter in deaf-blind children, who could not learn laughter through the senses. (Ruch & Ekman, 2001)) The fact that laughter is produced without using articulated sounds (see 1.4.1), moreover, supports its development before speech, and highlights its universal expressive-communicative function. (Ruch & Ekman, 2001)

But what is laughter? Any attempt to give an exhaustive definition of this term is likely to be vane. Indeed, there exist numerous descriptions of this phenomenon; if we search on dictionaries or encyclopedias, we can find different results:

"*...series of spasmodic and partly involuntary expirations with inarticulate vocalization, normally indicative of merriment, often a hysteric manifestation or a reflex result of tickling.*" (Dorland's Illustrated Medical Dictionary)

"*Spontaneous laughter is a motor reflex produced by the coordinated contraction of 15 facial muscles in a stereotyped pattern and accompanied by altered breathing. Electrical stimulation of the main lifting muscle of the upper lip, the zygomatic major, with currents of varying intensity produces facial expressions ranging from the faint smile through the broad grin to the contortions typical of explosive laughter.*" (Encyclopedia Britannica online)

“*An involuntary expression of merriment and pleasure; it includes the patterned motor responses as well as the inarticulate vocalization.* (Free medical dictionary online)

These definitions contain some common references, but on the whole they highlight different elements of laughter. This happens because of the very complex nature of laughter: there is no agreement even among authors whether laughter has to be considered as an emotion, a mechanical reflex, a communication mean, or other. Moreover, there are many different points of view under which it can be analyzed, as it interests many different disciplines, such as physiology, neurology, psychology, sociology, anthropology, etc..

Laughter is an essential signal in communication, inasmuch it conveys information about interlocutor's feelings, it eases social contacts, along with eliciting emotions in the others. In addition, more recently, the healthy effects of laughter have been demonstrated, such as its role against stress (see 1.2.2), and the utility of laughter therapy. (Urbain et al., 2010)

1.1.1 Different types of Laughter

Traditional argument of interest of philosophers, the study of laughter has not been a major theme of scientific research until last century, with a progressively increase of studies and publications on this subject, probably due also to the advent of new increasingly powerful techniques, such as functional magnetic resonance imaging (for a more detailed description of methods, see 1.3).

Very often, along with the impossibility of finding a universal definition, no agreement exists about how to classify different types of laughter. On the whole, whatever classification method is adopted, laughter remains an important universal cue for understanding emotions, and manifesting a certain emotion.

A typical general classification The most accepted classification often used by several authors (e.g. (Martin, 2007)) concerns the way laughter is generated and intentionally modulated:

- *Spontaneous*, emotional laughter: it is described as an impulse, it is hardly to suppress and control, causing a decrease of self-awareness and attention, and it is characterized by a clearly enjoyable feeling; this type of laughter can occur in response to humor, or others stimuli.
- *Voluntary*, contrived or faked laughter: people can laugh voluntary or on command, by producing a sound pattern like that of natural laughter, but experiencing a different degree of emotionality. This type of laughter can be considered unique in humans, since it is almost related to social rules (e.g. the polite laughter, when we laugh only to please a friend or the boss, but we didn't actually appreciate his joke).

Differences between the two types can be found in the larynx modulations occurring during laughter's production, as well as in the relative sound, the behavioural patterns (e.g. the presence of crow's feet, see 1.4.2) and the neural correlates.

A middle way between completely spontaneous and completely voluntary laughter could be considered when thinking about what some authors (e.g. Ruch)call "conversational laughter", that can include spontaneous laughter, but at the same time this is monitored according to social conversational rules. This typology occurs only in social interactions, during conversations, where laughter can perform different roles, even non emotional ones, such as speech enforcer, gap filler, and politeness signal. According to Foot and Chapman, “Laughter is used for maintaining the flow of interaction in our daily encounters: filling in pauses in our conversations and maintaining the interest and attention of our conversational partner. (...) We use laughter for these social purposes, much more frequently than we use it in response to actual humor stimuli” (Foot & Chapman, 1976). Moreover, social conversational laughter can be used in a conscious and goal-directed manner to influence and modify the attitudes and behaviors of social counterparts (Meyer et al., 2007).

Positive and negative laughter It is worth remembering that the different types of laughter are not always positive, but negative kinds of laughter are present as well in humans’ patterns. Their behavioural expression is often different from that of positive laughter types, such as hilarious laughter, also because their communicative message is different.

First of all, you can consider mocking laughter, i.e. laughter directed to a person, or a group, expressing contempt or ridicule. While this behavior can be viewed as positive referring to the person who laughs, since it causes amusement and enhance a sense of superiority towards the target of the laughing, this is not the case of the person who is laughed at. Moreover, there is empirical evidence that persons exist that are primarily concerned with *Gelotophobia* (Greek:*gelos*= laughter, *phobos* = fear), defined as the fear of being laughed at and appearing ridiculous to social partners (Ruch & Proyer, 2008; Proyer et al., 2008). Gelotophobes have difficulty in perceiving the difference between laughing *with* and laughing *at* (Platt & Forabosco, 2013). The main difference between these forms regards the fact that the latter typically involves a target, while the former does not exclude but includes the target person. Since gelotophobes often feel to be the target of laughing, they often experience laughter in a negative way. A large amount of studies about gelotophobia has been conducted by Ruch and its affiliates; a review of the state of the art in this field can be found in (Ruch et al., 2014).

In some cases laughter occurs in unlikely situations where to expect pleasant laughter: this is the case, for instance, of embarrassing and hysterical laughter. *a*) Embarrassing laughter can be considered as a coping technique, early learned by people as a behavior to face uncomfortable situations. Thus, one can use laughs during embarrassing situations in order to induce in himself the pleasant feelings which laughter is a natural expression of. Alternatively, you can resort to laughter to appear to others as though you are not upset but actually amused by the situation you are in, or to defuse others’ anger or blame, in case you are "caught in the act" (Morreal, 1983). Embarrassing laughter is often automatic, indeed an embarrassed person does not always think about all the possible effects of her laughter or consciously force a laugh in order to achieve them. For instance, insecure people in social interactions often express nervous laughter, but they are not aware about their laughing.

b) While embarrassing laughter can be voluntary, hysterical laughter belongs without doubts to the class of spontaneous laughter. It consists in an involuntary mechanism for reacting to a catastrophic event impossible to handle, both cognitively and emotionally, in a normal way. Often alternating with hysterical crying, this type of laughter is definitely not related to pleasure, but it is a reaction of the nervous system to reject the reality of the shocking situation, and face the horror through a behavior expressing amusement (Morreal, 1983).

The previous two examples of negative laughter referred to psychological responses. In other cases, laughter can be caused by disorders in the brain: this is the case of pathological laughter. This is often not related to a pleasant feeling, and the patient can be aware of the inappropriateness of its laughing, but powerless to control it.

Symptomatic laughter can derive from different causes, such as epileptic seizures, neurological disorders, lesions or tumors in certain areas of the brain. The most cited definition of the criteria defining pathological laughter is the one of Poeck, who describes it as arising: *a*) in response to non-specific stimuli; *b*) in the absence of a corresponding change in affect; *c*) in the absence of voluntary control of the extent or duration of the episode; *d*) in the absence of a corresponding change in mood lasting beyond the actual laughing (Poeck, 1969; Wild et al., 2003).

A more specific classification More recently, the increasingly use of databases has allowed subtler classifications of laughter. An interesting classification was carried out by researchers who attempted to classify the types of laughter using the clips found in the Belfast Naturalistic Database (Douglas-Cowie et al., 2003) and HUMAINE database (Douglas-Cowie et al., 2007). Starting with an initial categorization of 23 laughter types (Drack, Huber, & Ruch, 2009) they reached the laughter classification scheme in Figure 1.1. This classification is interesting because the descriptions in each category were developed in conjunction with users to ensure that they could be readily understood by non-experts (McKeown et al., 2012).

Expressing predominantly positive emotions	
Happy	Indicative of pleasure, contentment or joy because of a particular thing
Relieved	Laughter that signifies a concern or anxiety has been laid to rest
Thankful	Appreciative or expressing gratitude to a person
Hilarious	Unrestrained response to something that is simply found extremely funny
Giggling	Happy mixed with a sense that something is ridiculous or unimportant
Helpless	Laughter that is positive, but that the person has completely lost control of
Lustful	Salacious: expresses sexual arousal and anticipation
Mischievous	Playful but with the intent to case trouble
Boastful	Laughing out of self-admiration or excessive pride
Expressing predominantly negative emotions	
Angry	Laughter that conveys aggression or intimidation
Sarcastic	Laughter to convey that words spoken should be taken as cynical or mocking
Contemptuous	A laugh that expresses superiority to the person being laughed at, showing disdain or scorn towards them
Sullen	Laughter that indicates someone is being pressurised to behave in ways that he/she resents
Tense	Uncomfortable laughter. Used in situations where it is unsure what should be said
Embarrassed	A result of being self-conscious or an expression of confusion or shame
Hysterical	Laughter that the person has completely lost control of as a result of feeling that he/she has lost control of events.
Desperate	Frenzied laughter conveying a dire need for something
Sad	Laughter indicating regret that something has happened, with resignation that it cannot be changed
Expressing emotions with positive and negative elements	
Shy	Nervous and quiet, trying not to make feelings too obvious
Anxious	Experiencing unease and trying to lessen it with laughter
Apologetic	Awkward laugh used when an individual is trying to express an apology / show remorse.
Meaningful	A laugh to show there is more meaning to what has been said than is simply expressed
Cunning	Laughter that is shrewd, sly or deceitful
Taunting	Laughter directed at someone in particular. Intended to make fun of or belittle him/her.
Schadenfreude	Laughter expressing pleasure in the misfortune of another person
Other - not primarily expressing emotions	
Physical reflex	Response to physical prompt (usually tickling)
Surprised	A reaction to astonishment, when something happens suddenly or unexpectedly
Backchannelling laughter	Laughter that is part of a conversation, and conveys a routine acknowledgement of what the other speaker has just said
Polite laughter	Laughter aimed at being courteous and showing good manners
Contrived	A forced or planned laugh
Staged laughter	Completely forced. Usually detectable easily. The kind of laugh found in TV/films
Other	Not included in the list above

Figure 1.1: Classification scheme of laughter, outlined by (McKeown et al., 2012)

What will be investigated in this Thesis An analysis concerning all laughter typologies is out of the purpose of this Thesis, which however is oriented towards one particular type: hilarious laughter. This type of laughter occurs in response of a stimulus, such as a joke or a funny video, and does not require the presence of the others. It is without doubts a positive form of laughter, and it can be considered to belong to the class of spontaneous laughter. In particular, the evaluation study conducted in Paris concerned laughter triggered by funny stimulus while being inserted in a social environment, since the presence of a virtual agent(see 2.1 and 4). We could summarize that, by exploiting the presence of the virtual agent and its laughing, we were able to investigate how hilarious laughter can be affected by the interaction with a laughing agent.

1.1.2 Humor, Mirth and Laughter

With respect to this Thesis, we could say we will describe what Darwin highlighted in his works (Darwin, 1872), where he considered laughter as a way to communicate a particular emotion to the others, thus he did not consider different types of laughter, and defined the particular emotion communicated through laughter to be a pleasurable feeling closely related to joy. This emotion has been later labelled as "humour appreciation", "exhilaration", or "mirth". Thus, for our circumscribed analysis, we can adopt a similar point of view, and also the one of Martin (Martin, 2007), who defined laughter as an " emotional display expressing the positive emotion of mirth", and well distinguished the different roles of humour, mirth and laughter:

- humour as a **cognitive-linguistic** form of play (e.g. jokes and puns);

- Mirth as the **emotion** elicited by humour;

- Laughter as the typical **expression** of mirth.

It is interesting to notice that the original meaning of the word *humour* was quite different from the general meaning we now attribute to it, and then evolved several times towards its actual meaning. As MacGhee describes in its volume "humour: its Origin and development" (MacGhee & Pistolesi, 1979), the Latin word *humour* referred to the bodily fluids supposed to determine a person's general disposition. In particular, according to the temperamental theory born in ancient Greece and still present in medieval and Renaissance physiology, bodily moistures were yellow bile, associated with choler; black bile, associated with melancholy; blood, associated with sanguine temperament; and phlegm, associated with phlegmatic temperament. A normal person was characterized by a balance between the four humours, while an imbalance was thought to cause an alteration of the normal disposition. From this original definition, humour began to represent one's mood or state of mind, so that an imbalance was thought to be the cause of different forms of excessive behaviours, at which people used to laugh, calling the people with this behaviours as "humorists". As time passed by, the term has expanded its meaning, including anyone who was skilled at producing amusing or incongruous ideas or events.

Thus, here we can set the origin of the use of the term *humour* as an element triggering laughter.

1.2 Theories and functions of Laughter

Since a consensus does not exist among researchers about laughter definition, a dissent about its origin and functions is fairly predictable. Here we will describe the most important points of view under which laughter can be dealt with, while in the next paragraph more attention will be payed on theories about laughter functions.

1.2.1 Different perspectives about laughter's functions

Biological perspective Supported by Gruner (Gruner, 1978; Wilkins & Eisenbraun, 2002), this perspective highlights the biological roots of laughter, treating it as a universal expression, independent from speech, since it developed before it.

Social perspective Supported by Radcliffe-Brown (Radcliffe-Brown, 1952; Wilkins & Eisenbraun, 2002), this perspective collocates laughter in an interactive environment, and assumes it to be controlled by social cues, describing as a proof the fact that it is impossible to tickle oneself and evoke genuine laughter. Under this prospective, laughter has the function of maintaining social harmony and stability in the group.

Evolutionary perspective Supported, among the others, by McDougall (McDougall, 1922; Wilkins & Eisenbraun, 2002), it is based on the natural selection, thus it implies that laughter, since it still exists, must have a survival value. Hence, laughter is supposed to serve a bonding function, peacemaking function and a health-boosting function, acting as a "protective shield", triggering empathy for the others in order for the human race to survive.

When considering laughter's functions to have a survival value, there can be considered other different hypotheses compared to McDougall's. The main relevant is the point of view of some theorists, which suggest the role of laughter as a coping mechanism (e.g. (Berg & Van Brockern, 1995; Carlson & Peterson, 1995)), thus focusing more on the single individual, rather than on the group survival. Laughter as a coping mechanism allows a positive reframing of negative events, by replacing negative with positive affect. In this way one's ability to cope with negative states is increased, with positive effects in stress reduction, and, for example, physiologic benefits on patients (e.g. pain tolerance, diabetes, blood pressure).

The physiological benefits of laughter arise also from the physical act itself: it is comparable to mild aerobic exercise, hence it increases the release of serotonin, affecting positive mood. Comparing watching comedy video and running on a treadmill, it has been found that they produce the same increase in positive mood and decrease of emotional distress (Szabo, 2003).

1.2.2 Different theories for a common purpose

The theories and perspectives about functions of laughter and humour are legions, but they are not mutually exclusive: for example, the social function supported by

Radcliffe-Brown and others supporters of social perspective is very similar to the "protective shield" suggested by McDougall, or, as you can find below, the incongruity humour has been associated with release of energy, that is the basis of relief theory.

In the context of this Thesis we do not care about the possible functions of negative types of laughter, such as mocking or embarrassing laughter (see 1.1.1), as our interest focuses towards hilarious and social laughter, hence we will analyse in more details the theories that try to explain the functions of these types.

Relief Theory This theory associates mirth, and, consequently, laughter triggered by it, with the release of pent-up energy. Such release produces a sense of relief, followed by relaxation. Such release may be either cognitive, such as the release from anxiety, or a physical release of tension; both types were found to be involved in the reduction of diseases' symptoms. Indeed, since stress has physiological negative effects on health, and stress can be reduced by laughter, both in a physical and cognitive way, laughter can consequently help to reduce the negative effects of stress. Many studies support this point of view, for instance hilarious laughter triggered by funny videos has been found to reduce bronchial responsiveness to allergen (Kimata, 2004), as well as prevent exacerbation of diabetic neuropathy (Hayashi et al., 2007), and other positive effects on other different diseases.

Incongruity Theory This theory considers laughter as triggered by a violation of normal patterns, because of the presence of an incongruity close enough to be non-threatening, but different enough to be remarkable. The contradiction between expectations and experience is suggested as the cause of mirth and thus hilarious laughter. Incongruity humour allows the presentation of new perspectives; in this frame laughter can be viewed as a coping mechanism, inasmuch it occurs the reframing of a negative perception (the incongruity) in a positive one (the resolution of the incongruity and its appreciation). Moreover, this theory considers cognition as a fundamental element for laughter, as it allows the understanding of typical patterns and the resolution of the incongruities. This is supported by recent findings about the neural patterns underlying the finding of the incongruity and its resolution (see 1.5.2.1); moreover,the physiological benefits of laughter have been demonstrated even under the point of view of this theory, for instance Buchowsky found an increase of energy expenditure, and thus weight loss in long period, during incongruity laughter (Buchowski et al., 2007). This release of energy naturally linked the theory to the first one, well explaining how mirth can be generated.

Superiority Theory This theory differs from the previous ones, since it focuses on the societal functions of humour and laughter: they maintain social order by laughing towards those who refuse to comply with rules, and they reinforce group unity, by laughing at people outside the group, by eliciting similar emotions to the members of the group, increasing people's likeability and promoting affiliative and cooperative behaviours. While criticizing opposition, the members of the group experience a sense of supremacy over the other. Also in this case, the hypothesis of physiological benefits of laughter can be explained by considering that the bonding functions of laughter may contribute to maintain people's sense of well-being, inasmuch low social support is frequently associated with stress, depression and negative mood, which in turn are risk factors for diseases such as those related to heart and blood pressure. Superiority theory was already an argument of interest among great philosophers: Plato viewed laughter as a malice towards vice and powerless people, hence something irrational to avoid; Aristotle considered laughter as an expression of superiority towards the others, since we do not laugh at people placed at same level of ourselves.

On the whole, whatever point of view you prefer to follow, a common final point can be identified: hilarious laughter has positive influence on humans' organism, at both cognitive and physical level, and these positive effects can affect the single individuals as well as an entire group of people.

Finally, some authors highlight the role of humour and laughter in education, and suggest to introduce them into the educational process. According to coping perspective, laughter can help to face the stress of school, improving classroom climate, both for students and for teachers, thanks to a facilitation of communication, decrease of hostility and a general relaxed atmosphere (Carlson & Peterson, 1995). In this climate, learning process is improved: humour focuses the students' attention on the subject matter, and easier recall when using humorous associations has been demonstrated (Hauck & Thomas, 1972; Carlson & Peterson, 1995).

1.3 Methods to study laughter and humour

The interest on laughter and humour has belonged for years to the domain of philosophers, and, then to psychology. Only in 19th century the interest in laughter included the scientific world, initially with the development of neurologists' interest towards the study of pathological laughter (see 1.1.1), and, at the beginning of 2000's, with the investigation of normal laughter, in particular laughter triggered by humour. We can thus delineate a line of analysis that starts from philosophical and psychological investigation, passing through behavioural and physiological studies, coming to the brain as the primary subject of investigation. This line of analysis is in parallel with the development of technology and investigation techniques. Indeed, since in the past there was no instrumentation available for central nervous system (CNS) studies, the investigation of laughter was almost limited to the external effects, like facial movements, or diaphragm activation, and often their analyses based on the observations and annotations of experts. During the final quarter of 20th century much research has been conducted in the field of physiology, in particular to study the effects of mirthful laughter in human physiologic systems (Fry, 2002).

1.3.1 Facial expression analysis

As outlined before, the traditional methods based on direct observations by experts. Facial expressions can be objectively assessed, especially if the participants feel unobserved and do not voluntary regulate their expressions, thus they are an important element to take into account when assessing responses to humour stimuli.

While facial expressions related to laughter and expression of mirth will be described in 1.4.2, here we will highlight two methods to assess people's facial displays:

- Facial Electromyography (EMG): it is a technique that measures muscle activity by detecting and amplifying the tiny electrical impulses that are generated by muscle fibres when they contract (Wikipedia, 2014b). Since it uses electrodes, it is quite invasive and can inhibit spontaneous user responses. This tool if more economic than the following one, but at the same time it is less valid (Platt et al., 2012).

- Facial Action Coding System (FACS)(P Ekman, Friesen, & Hager, 2002): it is the leading coding schema that offers a reliable, valid and objective assessment of all visually discernible facial actions (Platt et al., 2012). It distinguishes 44 Action Units (AUs), i.e. the minimal units anatomically separate and visually distinguishable (Platt et al., 2012). This coding system can be used by certified FACS coders in conjunction to videos of participants, in order to objectively assess their facial expressions.

1.3.2 Methods in brain investigation

The interest on brain's role in laughter did not concern only its peripheral physiological effects, but aimed to investigate also its central role in perception, recognition and interpretation of humorous stimuli. This was reflected in the increased number of studies about localization of brain regions involved in these processes. Until the early 1990's the investigations of CNS would be obtained through three main procedures (Fry, 2002):

1. *Ablation cases*: they refer to traumatic loss of or deliberate (surgical) removal of CNS tissue (Brownell & Gardner, 1988; Fry, 2002).

   These events in some cases allow identifying functions of the damaged tissues, but they are not the best data source, inasmuch they present some disadvantages. For instance, the localization's accuracy can be affected because frequently a trauma can extend to surrounding tissues, thus it is difficult to distinguish what regions are actually involved in the function of interest, and what are not. Moreover, ablation cases cannot give information about a possible network involving the damaged area;

2. *Direct CNS tissues simulation*: this technique allows the localization of tissues participating directly in the functions active by the instrumental stimulation.

   In this case, as well as the previous one, no information is given about networks identification, and there exist also the risk that the stimulation effect diffuses through contiguous areas, thus it is difficult to know the range of experimental intervention impact (Fry, 2002);

3. *EEG*: ElectroEncephaloGraphy, that consists in the recording of electrical activity along the scalp.

   This method presents more advantages than the procedures described before, such as greater mobility and availability, higher resolution, absence of risks, and improved capacity for circuit identification. On the other side, its main disadvantages are the noise coming from nearby muscles or automatic movements like swallowing or blinking, and no differential of intensities is provided.

Only at the early 1990's the advance in technology allowed the substitution of traditional procedures with the use of new devices, such as fMRI, PET, TMS, and a consequent improvement of the quality and precision of information about brain functionality during laughter and experience of mirth. Thanks to these procedures, it is possible to investigate the dynamics of brain activities when reacting to stimuli, with little or none risks to the health, and having the possibility to study not only single cases of patients, but first of healthy people. In addition, results from these techniques disconfirmed the previous hypotheses coming form traditional studies, about the existence of a laughter center, and enabled the study of the immense complexity of of networks, continuous interactions and integrated structures involved in laughter and humour appreciation (Fry, 2002).

A pioneering fMRI study, whose results were confirmed by several lines of evidence using fMRI in healthy subjects (for a list of related studies, see (Wild et al., 2003)) was carried out by Goel and Dolan (Goel & Dolan, 2001). While the important findings of the study will be described in 1.5, it is interesting to report in this section a brief description about the procedure they used, since it could represent a method to study laughter and humour in a fMRI study. Subjects were presented with 30 semantic and 30 phonological jokes while undergoing fMRI scans. The participants were presented also with a corresponding baseline condition, where the punch lines were modified to be merely descriptive; they were also asked to make judgments about items funniness. Outside the scanner, they were asked to rate the jokes for funniness. If we want to generalize, an advisable procedure to follow while investigating neural correlates of laughter and humour should combine a presentation of the stimuli under the scan with a control group of stimuli in order to eliminate the common activated areas and find those involved in the specific processes under study; in addition, it is necessary to assess how the participants perceive the stimuli, to make sure about their actual funniness.

1.3.3 Experimental studies design and Humour

Another important field of investigation on laughter and humour concerns the conjunction of artificial intelligence and psychology. The contemporary computational approaches to humour can be improved by applying psychological experimental standards, that help to face typical issues such as individual differences in the appreciation of different kinds of humorous stimuli, and the lack of experimental control.

(Platt et al., 2012) well describe experimental standards in research on AI and humour when considering psychology; the rest of this paragraph will summarize their standards.

The typical purposes when conducting experiments in HCI and humour concern the evaluation of an interactive machine able to laugh and respond to human laughter, and the assessment of the impact of a virtual laughing companion on the humour experience of a human user. To measure the appreciation, the understanding and the cognition towards humour produced by AI, self-questionnaires are often employed, often using definitions and example items in order to reduce ambiguity in terminology interpretation. The most used scales of measure in this case are:

- Likert scale, where for each item is presented a scale of agreement/disagreement, generally using 5 or 7 scores;
- Semantic differential, where bipolar adjective pairs are used as the extremes of the ranking scale.

Methods and standards adopted in these experiments aimed to face different problems:

- Social desirability, i.e. the wish to comply with social rules or the experimenter: if participants understand what is being measured, in this case humour evaluation, there exists the risk for them to overestimate the funniness of the stimuli, since a good sense of humour is often considered as a highly desirable personal quality, and thus they want to be evaluated as good at judging the stimuli.

A possible solution to limit social desirability is the Cover Story. It is a false explanation given to the user during the instructions, with a credible context, without alluding to the true nature of the experiment. In this way, not only the participants' expectations are moderates, but also they are given something to focus, preventing effects of social desirability. A good cover story should give the idea that what is being measured does not relate to what actually is, the participants being then more likely to assess the stimuli at the true value.

- Confounding variables: the presence of some variables, out of the set of those manipulated by the experimenter (independent variables - IVs), can affect the measurements. In this case, a significant results could be due not to the IVs, but to another factor, for instance the gender, the age, etc...
- Influence of the experimenter: the interaction participant/experimenter can affect the mood states of the user, and could also alter the way in which the participant later engages with the virtual agent of humour stimuli, acting as a sort of confounding variable.
- Current mood of the user and its personal traits: certain individuals tend to generally appreciate, create, or laugh more easily and intensively at humorous stimuli than other do. Hence, some traits, like extraversion, can have effects on responses to humour, and then affect the generalization of the results of the study.

A possible solution for these three problems could be the Standardization: to limit the effect of confounding variables, it is fundamental to control the test environment sufficiently, following a specific pre-defined protocol on the procedure.

Standardization allows any replication of the study by any experimenter, thus a clear protocol should fix also the behaviours and engagement of the experimenter with the participant. If the experimenter acts in the same way with all the users, the error variance is highly reduced.

Finally, the problem of influence of personal traits and current mood could be faced by assessing them at the beginning of the experiment, using specific questionnaires: in this way it is possible to eliminate outliers, i.e. those participants with values high different from the average ones.

1.4 Behavioural pattern of laughter

As explained above the main interest of this Thesis is hilarious laughter, thus the following descriptions about behavioural patterns and neural correlates (see 1.5) will mainly refer implicitly to this category.

1.4.1 Acoustic features

Regarding the phonetic structure of laughter, the related literature contains a relative small number of investigations and data, and the presence of heterogeneous terminology, since different authors used different terms to define the same objects (Trouvain, 2003).

Despite these difficulties, a simple description of the acoustic features of laughter can be made. In the following paragraphs I will mainly refer to the terms used by (Ruch & Ekman, 2001), integrating them with some improvements suggested by (Trouvain, 2003).

It is not a case that in many languages the orthographic realization of laughter is "haha", "ahah" or "xaxa"; indeed these patterns perfectly reflect the typical laugh-pulse (see dopo), which consists of an alternating voiced-unvoiced pattern. Before describing this pattern and the acoustic segmentation of laughter into details, a general description of its time course is opportune.

According to Figure 1.2, the whole behavioural-acoustic events that, separated by inhalation, constitute a laughter episode, are called laughter bouts.

Each bout may be subdivided into three steps:

1. Onset: this short and steep part, especially in explosive laughter, is the pre-vocal facial pattern;

2. Apex: it is the proper vocalization part of the bout, and it is composed by laugh cycles, i.e. the common "hahaha". According to Trouvain's segmentation, we can consider laugh cycles under two deeper levels:

   (a) at a syllabic level, laugh cycles are sets of repetitive vocalic segments, called interpulse intervals by Ekman and Ruch. These syllabic nuclei are interspersed with pauses, and their maximum number in a cycle is limited by the lung volume.

   (b) at a segmental level, the laugh-pulse is constituted by

      - A consonant segment, in particular an aspirated h typed sound, considered as an inter-pulsed pause by some authors (e.g. Ekman and Ruch) while by others (e.g. Provine) it is considered as a proper sound. These different interpretations are related to the decision about considering proper sounds only those produced by vibration of vocal cords, or if also an aspiration such as *h* can be considered as a type of sound. According to the different points of view about the actual nature of this consonant segment, also the denomination of the laugh-pulse changes: on one hand, it is described as segment-pause alternation (h as a non-sound), one the other hand it is considered a segment-segment alternation (h as a sound).

- A vowel-like utterance nucleus, the sound pulse of the alternation, whose duration is less than the consonant segment's; the particular vowel produced is supposed to be the neutral vowel schwa, achievable by opening the mouth and lowering the jaw, without needing any other articulator.

3. Offset: the post-vocalization part, characterized by a long-lasting smile, and its smoothed fading out.

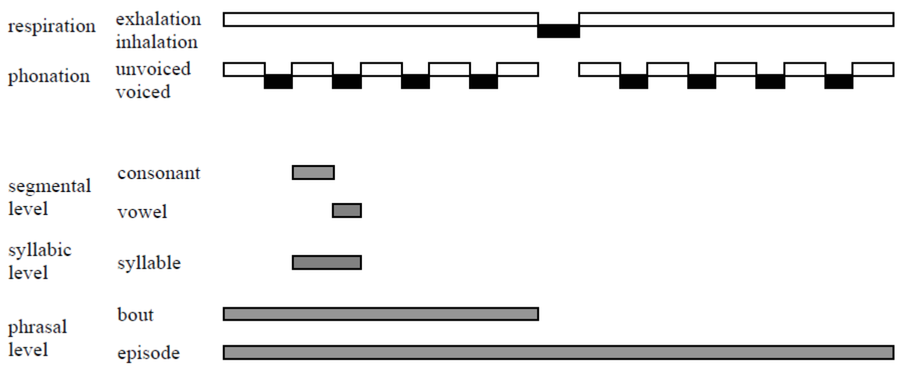


Figure 1.2: The time course of laughter. (Trouvain, 2003)

Other elements are worthy of being described when analysing the acoustic realization of laughing, since their fundamental role in sound production and modulation.

Laughter Respiration The respiration process during laughter is different from the typical one, involving additional muscles or modifying the natural respiration patterns. The first difference with a typical respiration cycle is that laughter begins with a forced exhalation, instead of an inspiration; this initial forced exhalation brings the lung volume down to around functional residual capacity (i.e. the volume that remains after a normal expiration). Then, a rhythmic pattern of repeated expirations of high frequency and low amplitude occurs, which allows the phonation of "hahaha" of the laugh cycle. Inhalation is expected to interpose between expiratory phases, in case of sustained laughter. At the end of the laughter bout the lungs reach residual volume (i.e., the air volume remaining in the lungs after maximum expiration). The muscles involved with saccadic contractions during laughter are typically passive during normal expiration; these auxiliary muscles are, for example, the diaphragm, the abdominal and the rib cage muscles.

The modulation of the air stream is modulated not only by respiratory muscles, but an important work is made by the larynx and other supra-laryngeal articulators, which co-determine the phonation of laughter.

Thanks to the initial forced exhalation, which makes the trans-diaphragmatic air pressure increase, the air stream ascending from the lungs can reach the larynx; in turn, thanks to the rapid and continuous adjustments of the latter, which affect the opening and closing of the glottis, the air stream is interrupted, provoking the vibration of vocal cords. The modulation of the vibrations trough the vocal tract's particular shape produces the final sound, and then the air goes out through the mouth or the nose.

Laryngeal modulation and phonation Laryngeal adjustments produce the vibratory cycles of vocal cord, characterized by phases of opening and closing of the glottis. The frequency of vibratory cycles per laugh-pulse was estimated by Moore and Van Leden to be around 5-15. Different vibratory patterns, in terms of duration and rates, distinguish different types of laughter. Ekman and Ruch (Ruch & Ekman, 2001) divided the laryngeal adjustments during vibratory cycles into four stages:

1. *Interpulse pause*: it is the phase where the *h* sound is produced. During it, the authors defined as "an instant of quiet", the arytenoid cartilages are open, allowing the air flow through the larynx, thus the vocal cords are motionless and no sound is produced, except for the aspirate sound of h.
2. *Adduction*: it is the stage of closing glottis. The arytenoid cartilages carry the cords toward each other, so that the glottal space becomes increasingly thin, allowing the vocal cords to start vibration.
3. *Vibration of vocal cords*: in this phase, where the cords continue to vibrate, the glottis is closed. The vibration of the cords allows the production of the vocal sound involved in the laugh-pulse.
4. *Abduction*: during this stage the arytenoids and vocal cords move in their rest positions, allowing the opening of the glottis and the diminution of vibration.

Supralaryngeal modulation Finally, the last important modulators, whose function is to produce the final acoustic output from the original sound coming from the glottis, are the articulators of the resonance tract. Some of them are:

- The lips, whose protraction or retraction cause variations in the pitch;
- The jaw, whose lowering, by modulating the degree of aperture of the mouth, affects the sounds;
- The cheeks, which, in addition to be part of facial actions, affects the sound, since their contraction change the mouth aperture and tenses skin of the upper ending of the vocal tract.

The tongue, however, while having an important role in speech production, seems to stay in a resting state position during laughter.

1.4.2 Facial Expressions and Body Movements

Everyone can describe laughter manifestation as involving the whole organism, in an uncontrollable and compulsive way, contagious and difficult to fake. This action pattern, similar to a reflex but on the other hand dependent on emotional states and social context, is maybe one of the most difficult to study, since the difficulty to reproduce natural laughter under command and in experimental settings.

During centuries, laughter manifestation analysis has focused on different components, according to the specific authors. For example, Dearborn (Dearborn, 1900; Askenasy, 1987) considered laughter as a general reflex, highlighting the role of diaphragm neck and hand movements, as well as Koestler , who suggested that laughter is the only domain in which a highly complex mental stimulus as humour produces a stereotyped reflexive response (Martin, 2007), while more recently the term reflex was refused, and laughter was linked to motivational and emotional states and communication functions (Martin, 2007); according to other authors, the main important component was considered to be the facial expression (Chevalier-Skolnikoff, 1973; Pollio et al., 1972; Askenasy, 1987). According to Askenasy (Askenasy, 1987), laughter's manifestation combines both facial expressions and general reflex movements, along with the major component related to breathing motor activity, i.e. the abrupt expiration due to a sudden contraction of the intercostal muscles, followed by the saccadic cycles described in 1.4.1.

Even the difficulties in measuring facial display of laughter, researchers agree on some universal elements that distinguish laughter from other emotional expressions. First, its connection with smile is evident, being the latter present at the beginning of a laughter episode, and smoothly appearing again at the end of laughter (Pollio et al., 1972; Martin, 2007). The only problem is to state if there exists a typical type of smile connected to laughter. Ekman and his colleagues, among the 18 types of smiles that they classified, have identified the only one associated with genuine enjoyment of amusement: the Duchenne display, named from the French anatomist who first distinguished this pattern from different kinds of smiling (1862). The two muscles found to be activated in almost the empirical studies, despite the different hypotheses and different methods of investigation used, are:

- The zygomatic major, a muscle in the cheeks whose contraction pulls the lip corners backwards and upwards;
- The orbicularis oculi, a muscle surrounding the eye socket, which, by raising the cheeks, causes the famous "crow's feet", i.e. the skin wrinkling at the outer side of the eyes.

This last muscle, which is less subject to voluntary control, is activated only in in genuine enjoyment smiles, thus it is a fundamental element to recognize genuine laughter. These contractions producing this display are symmetrical and synchronous.

Since the presence of Duchenne smile is not sufficient to define laughter's facial display, the latter must be characterized by additional components not present during smiling: these often includes the simultaneous activation of additional muscles, such as those involved in opening the mouth and lowering the jaw (by relaxation of masseter and pytergoids). Further facial changes often described are a brightening of the eyes, that sometimes produces tears, and a reddening of the face, due to vasodilation.

In Figure 1.3 a summary of the main facial display, by Askenazy, according to previous descriptions of Dearborn and other more recent authors (Askenasy, 1987; Dearborn, 1900). Along with other secondary components, Duchenne display is present in the figure, in terms of "elevation of the cheecks", "raising the corner of the eyes" (Orbicularis oculi), and the representation of the lips pulled upwards (Zigomatic major).

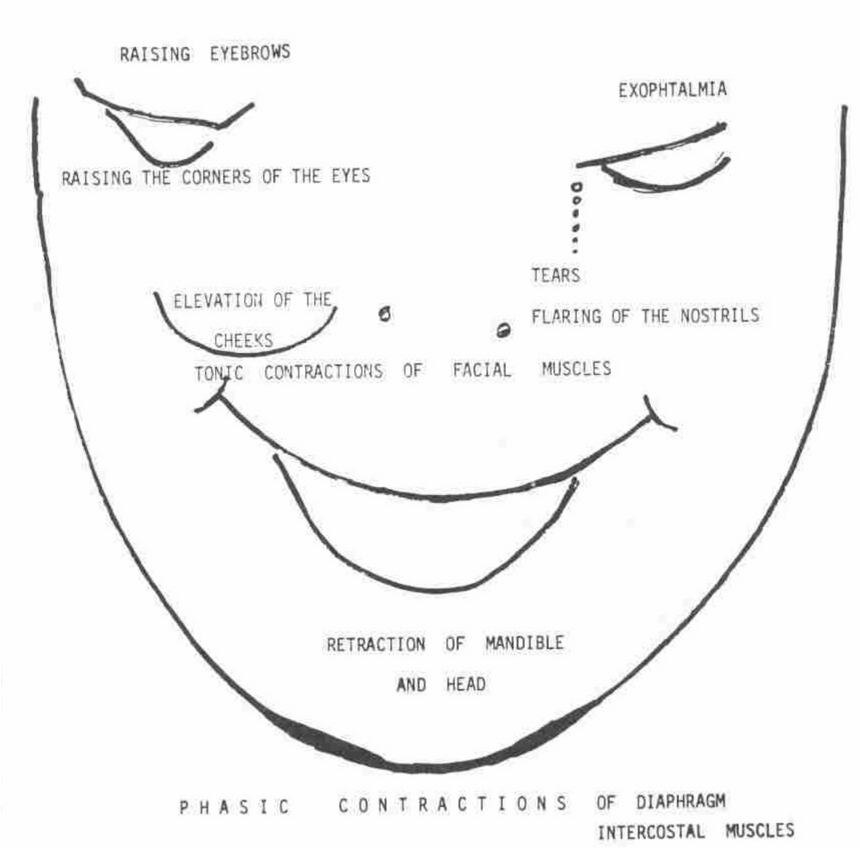


Figure 1.3: Scheme of facial expression of laughter. (Askenasy, 1987)

Since laughter, as Darwin stated (Darwin, 1872), is essentially an emotional expression, used to communicate to others a pleasurable feeling closely related to joy, i.e. humour appreciation, or mirth, its expressive display is related to the intensity of the emotion. Thus, at low levels of intensity, mirth can be expressed by only one weak smile, then, as soon as the intensity becomes stronger, the expression of it passes from an increasingly broad smile, to audible laughter, to an involvement of the whole body. The main actions occurring during exaggerated laughing are coupled with respiration, for example the backward tilt of the head facilitates the forced exhalations, while forced inhalation between cycles raises and straightens the trunk. Moreover, shaking of the shoulders and vibration of the trunk have been described (Ruch & Ekman, 2001; Darwin, 1872).

1.5 Neural correlates

The exact brain mechanisms underlying laughter are not fully understood. Based on evidence from cases of pathological laughter, instead of laughing centre, it seems to exist a system of networks of cortical and subcortical circuits, being motor areas a substantial part of them (Meyer et al., 2007). Only recently neuroscientists are beginning to piece together the circuits of the brain that are involved in the positive emotion of mirth and the production of laughter, although many of the details are still unknown (Wild et al., 2003; Martin, 2007) What is shown by the results of different studies of the last decades is that laughter can be described according to two aspects (Fry, 2002):

- A motor component, concerning the behavioural patterns, such as the bilateral facial and bodily movements, and their related neural correlates;

- A cognitive emotional component that includes the different aspects of what is called humour or mirth (using the nomenclature of adopted in this Thesis, the emotional state behind hilarious laughter).

Thus, according to Martin (Martin, 2007), there exists a strong relation among humour, mirth and laughter, so that the latter can be viewed as “a cognitive-linguistic form of play that elicits the emotion of mirth, which, in turn, is typically expressed through laughter” (Martin, 2007).

Since the fundamental dual role of the brain (Fry, 2002), which by itself on one hand represents the central organ of perception and interpretation of humour (cognitive component of laughter) and at the same time participates in its peripheral physiological effects (motor component), the areas associated with these two functions are supposed to be separated.

1.5.1 Neural basis of laughter motor patterns

While analysing the exterior behaviours of typical laughter, a further distinction can be made. Many authors support the existence of two independent neuronal pathways associated with laughter production: a voluntary pathway and an involuntary one (Martin, 2007; Wild et al., 2003).

A strong evidence of this dissociation comes from patients with volitional facial paresis who are able to produce natural hilarious laughter, and other patients who, on the other hand, are unable to express facial expressions when they are amused, but they can smile voluntary on command (Martin, 2007; Wild et al., 2003).

The fact that these networks include different regions may imply a difference also in the external behaviour. Probably this difference is almost imperceptible, but we can remember for example the interesting presence of "crow’s feet" only during real enjoyment, while it is well-high impossible to reproduce it voluntarily 1.4.2.

The two independent neural pathways are:

- Voluntary: this pathway concerns the unemotional laughter, thus not related to a feeling of mirth. Voluntary laughter includes the exterior pattern that you can produce under command or when you simulate laughter.

In this network, inputs origin in the motor cortex, in the frontal lobe (Brodmann area 4),and, through the upper motor fibres' aggregations of pyramidal motor system, they directly reach the face thanks to lower motor neurons of cranial nerves. Fibres of this pathway reach the ventral areas of mesencephalon and pons in the brainstem.

- Involuntary: this network is related to the emotional laughter, thus it is implicated in the realization of patterns during hilarious laughter. In this case, the origin of the inputs is assumed to be in the subcortical nuclei, instead of the motor cortex, then the signal reaches the face in an indirect way through the extrapyramidal system, the neural network of motor system that causes involuntary reflexes. The regions involved in the emotional pathway are indeed emotion related areas, such as basal ganglia, limbic system and brainstem. In particular, conversely with respect to the voluntary network, where signal proceeds ventrally, here the dorsal tegmental areas of mesencephalon and pons are involved.

In other words, when you voluntary produce laughter, you are not emotionally involved, but you are focusing on motor actions; this is indeed reflected in the activated network, where the inputs originates from the motor areas. During emotional laughter, instead, the behavioural actions are the consequences of the perception of humour, which activates emotion induction areas, which in turn affect emotion effector sites, including the motor and premotor areas of the cerebral cortex (for facial and bodily movements), the hypothalamus (for autonomic responses), and other areas such as PAG and RF. (Wild et al., 2003) propose that, while voluntary laughter is initiated in the motor cortex, through a direct excitation of these areas, during emotionally driven laughter there occurs an inhibition of the inhibitory system that monitors them. Such inhibition has been hypothesised also in pathological laughter.

Despite claiming the existence of a functional division, (Wild et al., 2003) at the end of their description postulate the presence of a final common pathway for laughter, located in the brainstem (dorsal upper pons), that coordinates the respiratory, laryngeal and facial components of laughter described in (par behavioural). This laughter-coordinating centre is connected to the PAG and the RF, which are an interface for behavioural control (Neurology, 2012).

The authors represent their idea of laughter network in Figure **??**.

1.5.2 Neural basis of hilarious laughter

Since the main argument of the Thesis is hilarious laughter, we will investigate into details the neural substrates of this type of laughter. It is important to keep in mind that, even though this will not be further analysed in this Thesis, the behavioural patterns of laughter can be triggered not only by humour and mirth, but there exist others causes, more mechanical rather than cognitive. For example, when exposed to tickling or laughing gas, people can express laughter without experiencing a sense of mirth.

While also animals have been found to show laugh-similar patterns (e.g. play face in monkeys), laughter as expression of a feeling of amusement and hilarity is unique

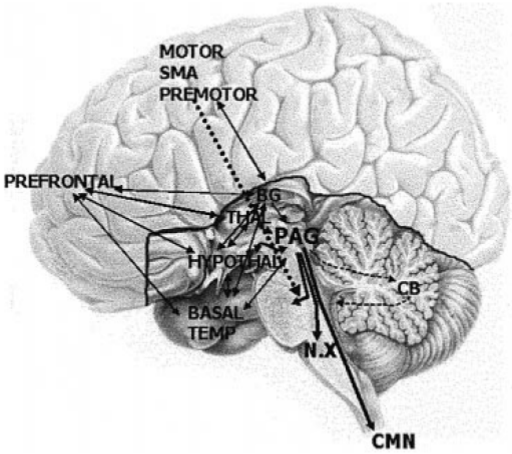


Figure 1.4: A proposal for laughter network. (Wild et al., 2003).
BASAL TEMP= basal temporal lobe including amygdala; CB= cerebellum; CMN= cervical motor neuraons; BG= basal ganglia; HYPOTHAL= hypothalamus; MOTOR= motor area; N.X= vagal nerve nucleus; PREFRONTAL= medial and dorsolateral prefrontal cortex; PREMOTOR= premotor area; THAL= thalamus.

of humans, and it has been discovered to include two components, anatomically separated, whose correct functioning allows a joke or a pun to be successful: a cognitive processing is necessary for humour detection, but not sufficient; indeed it should be followed by an affective feeling of amusement, to fully appreciate the joke.

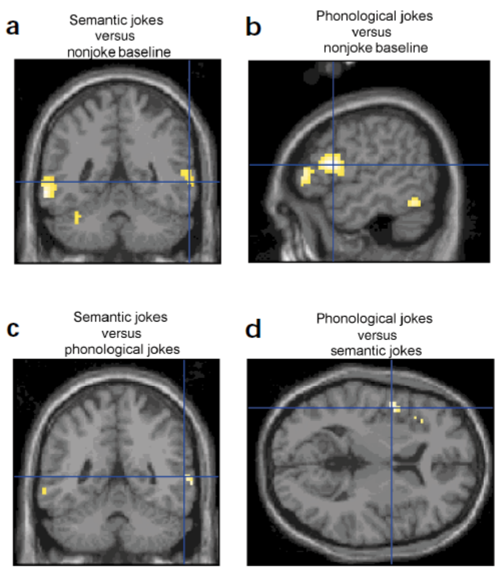
1.5.2.1 Cognitive components

The first step to appreciate a joke or a pun concerns the detection of the elements that contribute to the underlying funniness. This detection obviously requires a cognitive processing to be performed. In turn, this cognitive processing has been found to be type-specific, reflecting the incongruity-resolution theory (see 1.2.2) a semantic juxtaposition allows the perception of an incongruity between stored expectations and a punch line, while this incongruity is resolved by a phonological processing. Moreover, these two different activities involve different brain regions.

Semantic Processing While isolating the mechanisms underlying juxtaposition of mental sets, Goel and Dolan directly compared the semantic and phonological joke conditions. The areas activating during semantic jokes but not during the corresponding non-joke baseline were found to outline a bilateral temporal lobe network, in particular the left and right posterior middle and left posterior inferior temporal gyri (BA 21/37), involved in semantic processing of language, as well as processing and maintaining alternatives and integrative processing requiring global coherence.

Phonological Processing When Goel and Dolan compared brain activation during phonological jokes with respect to the corresponding non-joke baseline, they found a left hemisphere network, centred around speech sounds processing regions, in particular the left inferior frontal gyrus (BA 20/37) and left insula / pre-central gyrus.

Since humorous laughter is unique in humans because it is different from other mechanical-triggered laughters that can be found also in animals, the cognitive components of this type of laughter could be related with language regions, another characteristic unique in humans. According to this hypothesis, during hilarious laughter areas related to language should be activated. Indeed, the semantic and phonological networks detected by Goel and Dolan (see paragraph above) support this hypothesis; moreover this relation between language and mirth was confirmed by Yamao et al. (Yamao et al., 2014), who studied neural correlates using direct electrical cortical stimulation: they observed laughter and mirth in patients while stimulating left basal temporal area, left inferior frontal gyrus and left medial frontal lobe. These areas all belong to the left hemisphere, which is well-known to be dominant for language, thus their findings suggest a relation between left hemisphere and mirth, even though the authors warn to interpret them carefully.



An anatomical dissociation occurred in the cognitive processing of semantic and phonological jokes, whereas a common region underlay affective processing. (**a**) Processing of semantic jokes activated left posterior middle temporal gyrus (BA 21/37; –60, –54, 0; Z = 4.65), left posterior inferior temporal gyrus (–63, –51, –9; Z = 4.81), right posterior middle temporal gyrus (BA 21; 63, –48, 0; Z = 3.21) and cerebellum (36, –63, –30; Z = 4.81 and –36, –54, –33; Z = 4.68). (**b**) Processing of phonological jokes activated left posterior inferior temporal gyrus (BA 37; –60, –54, –12; Z = 5.34 and –48, –57, –21; Z = 5.08), and left inferior frontal gyrus (BA 44, 45; –51, 15, 15; Z = 5.32 and –54, 36, 6; Z = 4.93). (**c**) A direct comparison of semantic jokes with phonological jokes revealed activation in right posterior middle temporal gyrus (BA 21/37; 66, –45, 0; Z = 5.01 and 60, –48, 6; Z = 4.84) and left posterior inferior temporal gyrus (BA 20/37; –63, –39, –15; Z = 5.19). (**d**) The reverse comparison revealed activation of left inferior precentral gyrus/insula (BA 44/4; –48, 0, 6; Z = 4.71).

Figure 1.5: From (Goel & Dolan, 2001)

1.5.2.2 Affective components

Detecting the humorous components underlying a joke or a pun is not sufficient to trigger hilarious laughter. In fact, you need to find this incongruity as funny, to feel a sense of mirth. Goel and Dolan investigated this affective component by asking the participants to rate the jokes and puns for funniness, in order to compare the activations associated with jokes reported as funny with those associated to jokes reported as not funny. The affective appreciation was found to be modality-independent, thus the same for both semantic and phonological processing, and its activity involved the bilateral cerebellum and the medial ventral prefrontal cortex.

The involvement of the mesolimbic dopaminergic reward system which MVPFC belongs to could explain the positive feeling associated with mirth.

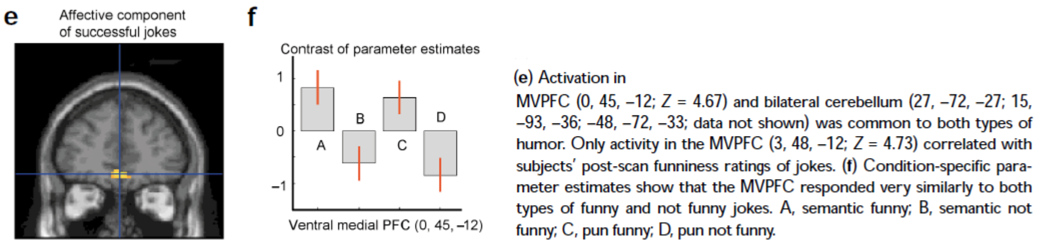


Figure 1.6: From (Goel & Dolan, 2001)

Chapter 2

Virtual Agents

This chapter provides an overview about virtual agents: their definition and types are given in 2.1, while a common framework proposed for these systems is illustrated in 2.2, then the positive effects of integrating laughter in virtual agents are described in an overview of laughter theories is given in 2.3, a European project with that intent is presented in 2.4, and finally some recent works focusing on laughing agents are reviewed in 2.5.

2.1 Definition and types of virtual agents

In artificial intelligence, an intelligent agent (IA) is an autonomous entity which observes through sensors and acts upon an environment using actuators (i.e. it is an agent) and directs its activity towards achieving goals (i.e. it is "rational", intelligent). Intelligent agents may also learn or use knowledge to achieve their goals. They may be very simple or very complex: a reflex machine such as a thermostat is an intelligent agent, as is a human being, as is a community of human beings working together towards a goal (Wikipedia, 2015d).

Different types of virtual agents exist, according to their level of autonomy, interaction with the user, and other characteristics. Embodied Conversational Agents are those towards which research is increasingly focusing, and they are used in several domains; we will describe their main properties, possible applications and evaluation parameters, then we will briefly describe other typologies of agents, less sophisticated but very used in the past and also at present in some specific environments.

2.1.1 Embodied Conversational Agents

An Embodied Conversational Agent (ECA) is a computer-generated animated character that is able to carry on natural human-like communication with users. (Urbain et al., 2009). According to their name, ECAs are:

- **Embodied**, thus a personification of the machine, in the form of a physical body situated in the virtual environment, or even only an animated talking head;

- **Conversational**, thus interactive, social, capable of engaging in conversations with one another or with humans employing the same multi-modal interaction means (Natural Language Processing, non-verbal behaviours, etc...), that humans do;

- **Agent**, since they maintain the characteristics of virtual agents, such as rationality, proactivity, autonomy.

ECAs can be developed in different environments, such as smart objects, virtual reality, or the computer; they are modelled from human data gathered from both annotated corpora and related works like psychological knowledge.

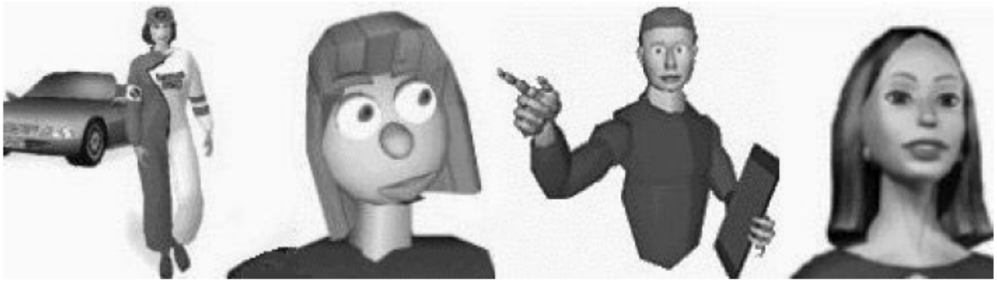


Figure 2.1: Examples of ECAs. From left to right we see: Jennifer James, a car saleswoman designed by Extempo Systems Inc. who attempts to build relationships of affection, trust and loyalty with her customers; Karin, informing about theatre performances and selling tickets (Heylen et al., 2001); Steve, educating a student about maintaining complex machinery (Johnson & Rickel, 1997); and Linda, a learning guide (Slater, 2000).

Starting from Computer Are social Acts (CASA) paradigm (Nijholt, 2002; Reeves & Nass, 1996), that states that humans respond to computers as if they were social entities, thus attributing attitudes and personality traits to them, as well as behaving following social conventions, and reacting to them as they would react to a human, ECAs would be a more powerful way for humans to interact with their computers since they use a primary and early-learned skill of humans, i.e. conversation, that is a very defining characteristic of humanness and humans' interactions (Cassell et al., 2001). Face-to-face conversation applied to human-computer interfaces is often taken into consideration by designers, indeed this type of interaction allows for much richer communication, and enables pragmatic acts such as turn taking. Moreover, users have been found to prefer non-verbal visual indication of an embodied system's internal state to a verbal indication (Wikipedia, 2014a; Marsi & van Rooden, 2007).

2.1.1.1 Properties of ECAs

Cassell listed a series of properties of humans in face-to-face conversations that ECAs should acquire (Cassell et al., 2001):

- The ability to recognize and respond to verbal and non-verbal input;

- The ability to generate verbal and non-verbal output;

- The ability to deal with conversational functions such as turn taking, feedback, and repair mechanisms;

- The ability to give signals that indicate the state of the conversation, as well as to contribute new propositions to the discourse.

In addition, she defined other fundamental conversational properties while building an ECAs. The most important thing for her was to think about functions rather than behaviours: during conversations, behaviours are only surface features of higher level elements, since they have a role or a function in the exchange. The same behaviour can have different communicative effects, according to situational factors, as well as one communicative function can be expressed through different types of behaviours. Hence, an ECA should be a system operating on functions, dealing with input and outputs.

Another important property concerns the fact that humans assume that behaviours occurring in synchronization carry the same meaning or the same function, thus this could be exploited by the agent while deciding the actions to perform.

The agent should also distinguish between propositional and informational information. The first one achieves the goal to shape the actual content of the conversation, and includes meaningful speech and non-verbal behaviour useful to complement the content of the speech; the latter has the goal of creating and maintaining the communication, through the use of cues that regulate the process, such as non-verbal behaviours and regulatory speech. Also in this case the role of functions is fundamental, since every behaviour is performed with a purpose.

Finally, entrainment is considered an important characteristic as well: through non-verbal cues such as gaze, eyebrow raises and head nods, the conversational partners synchronize their turns and behaviours, and this increasingly occurs during the conversation.

Cassell tried to implement the properties described above in the agent Rea, whose domain of expertise is real estate and she has access to a database of available condominiums and houses for sale in Boston. She can display pictures of those properties and of their various rooms, and point out and discuss their salient features.

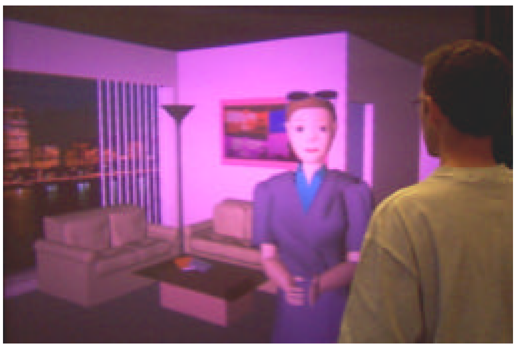


Figure 2.2: User interacting with Rea. (Cassell et al., 2001)

2.1.1.2 Applications of ECAs

The specific characteristics of ECAs described before make them suitable for several different contexts and roles (Pauchet & Sabouret, 2012):

- Assistant: this type of agent has the function of welcoming and assisting the user in understanding and using an application or a website. It has the capacity to resolve help requests (often inputs in natural language or speech) issuing from users with poor knowledge about a component (application, website, etc.).

  Assistant agents are often present in e-commerce, or in websites which offer assistance to customers.

- Tutor: well-built ECAs can be useful for students in the context of e-learning, but also for patients, in monitoring systems, both for psychological and physiological problems.

- Partner: this role is thought for actors in virtual environments such games, virtual communities, working groups. The ECA in this context can help for example to solve a problem cooperatively or to moderate a virtual meeting or a teleconference.

- Companion: ECAs can be used in some cases without specific functions, but only to become a virtual friend in long term relations.

2.1.1.3 Evaluation of ECAs

Evaluation of performances is an important step when building an ECA, in order to find what works and what can be improved. Generally, the evaluation of ECAs assesses several properties, both objective and subjective (Pauchet & Sabouret, 2012).

- Objective Evaluation:
  - Efficiency: it assesses the actual performance of the couple user-agent when carrying out a given task.
  - Usability: it measures the capacity of the user to have a good comprehension of the functioning of the system and the fluency of the control.

- Subjective Evaluation:
  - Friendliness: it assesses the "feeling" of the user about the features of the system (attraction, commitment, aesthetic, comfort).
  - Believability: it measures the "feeling" of the user about the fact that the agent can understand the user's problems and has the capacity to help with real competence.
  - Trust: it assesses the "feeling" of the user about the fact that the agent behaves as a trustable and cooperative entity.

2.1.2 Chatbots

The term derives from *chatterbox* and *robot*: a chatbot is a program (in form of artificial person, animal or other creature) capable of Natural Language dialogue, usually exploiting key-word spotting. The dialogue could be a text based (typed) conversation, a spoken conversation or even a non-verbal conversation. Chatbots can run on local computers and phones, though most of the time it is accessed through the internet (Wikipedia, 2015b).

The goal of this type of agent is only to chat, thus the main difference with ECAs is it does not require embodiment.

Figure 2.3: Some examples of chatbots. Their graphic design can vary from very simple ones to sophisticated human-like characters. (Chatbots.org, 2015a)

A chatbot can be interesting, inspiring and intriguing. It appears everywhere, from old ancient HTML pages to modern advanced social networking websites, and from standard computers to fashionable smart mobile devices (Chatbots.org, 2015a).

Chatbots talk in almost every major language. Their language skills vary from extremely poor to very intelligent, helpful and funny. The same counts for their graphic design: you can find a cartoonish character like one drawn by a child, but also photo-realistic 3D animated characters, which are hard to distinguish from humans (Chatbots.org, 2015a). They are often static, i.e. they are a still image.

The first chatbot is generally recognized to be ELIZA, a conversational artificial intelligence therapist designed in 1966 by Joseph Weizenbaum. The main mechanism of ELIZA consisted in rephrasing many of the patient's statements as questions and posing them to the patient. It was meant to be a simulation of a Rogerian psychotherapist. Using almost no information about human thought or emotion, ELIZA sometimes provided a startlingly human-like interaction, event though it often risked to create a conversation without sense; an example can be read in Figure 2.4.

ELIZA was implemented using simple pattern matching techniques, but was taken seriously by several of its users, even after Weizenbaum explained to them how it worked (Chatbots.org, 2015b).

2.1.3 Avatars

Avatars are graphical entities dedicated to the mediating process between humans in distributed virtual environments (Pauchet & Sabouret, 2012).

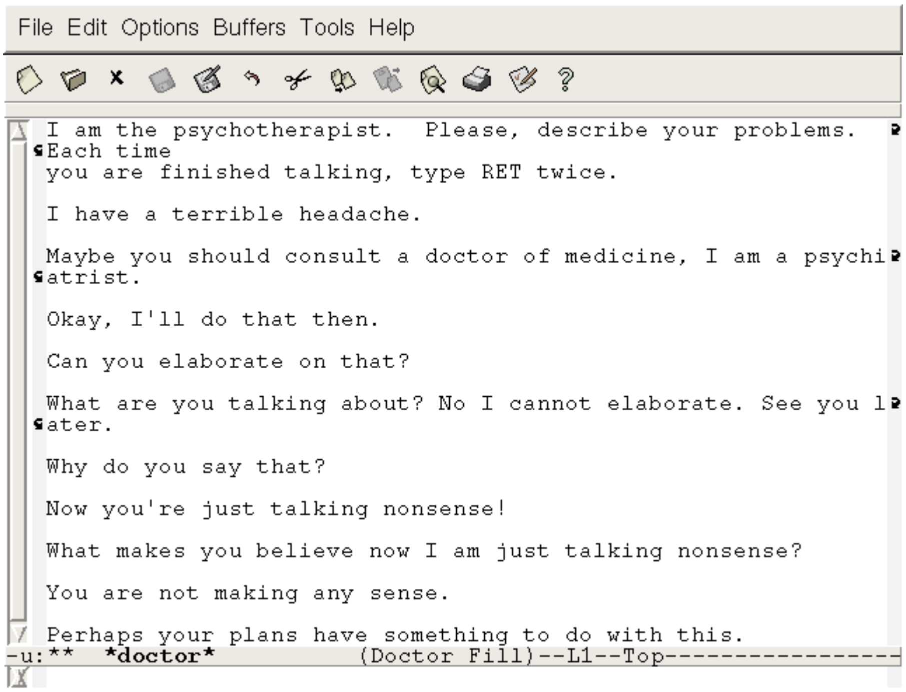


Figure 2.4: An example of conversation with Eliza. (Wikipedia, 2015c)

The main difference between Avatars and ECAs concerns the level of autonomy: while Avatars are a personification of a real human who controls them, ECAs are personifications of a computer program, thus they are artificial agents, completely autonomous from users' commands.

A famous example of applications using Avatars is Second Life, a web virtual simulation of the social life of ordinary people, developed by San Francisco-based Linden Lab and launched in 2003. Second Life users (also called Residents) can create virtual representations of themselves, and are able to interact with other avatars, places or objects. They can explore the world, meet other residents, socialize, participate in individual and group activities, build, create, shop and trade virtual property and services with one another (Wikipedia, 2015a).

2.1.3.1 Mixed Communities

There can exist computerized environments that integrate humans and artificial agents: in these environments, called Mixed Communities, virtual characters driven by people in a virtual world, i.e., avatars, interact with ECAs. These communities can help in virtual training, purchasing, chatting, brainstorming, decision groups, and other domains.

2.1.4 Non Player Characters

In the context of games, a specific type of virtual agent is often present in the background: a non-player character (NPC), sometimes known as a non-person character or non-playable character, is any character that is not controlled by a

Figure 2.5: An example of Second Life environment.

player. In video games, this usually means a character controlled by the computer through artificial intelligence. In traditional tabletop role-playing games the term applies to characters controlled by the gamemaster or referee, rather than another player. NPC behaviour in computer games is usually scripted and automatic, triggered by certain actions or dialogue with the player characters (Wikipedia, 2015e).

2.2 SAIBA framework for ECAs

In order to stimulate human-like multi-modal communicative behaviour, advanced ECA systems need to incorporate a whole range of complex processing steps, from intent to behaviour planning to behaviour realization including some sort of scene or story generation, multi-modal natural language generation, speech, synthesis, the temporal alignment of verbal and non-verbal behaviours, and behaviour realization employing particular animation libraries and engines.

In the last decade a significant number of ECA systems and application have been developed, and in parallel an almost equal number of usually XML-based markup and representation languages (Krenn et al., 2011). Different types of languages were used to represent multi-modal communicative behaviours:

- Markup languages: for annotating text with behaviour directives. They define sets of markups that give the non-expert user (usually a web designer) the possibility to annotate text with high-level behavioural information in order to produce pre-scripted presentations for ECAs. They are indispensable in application development, because the application designer need not to necessary be an expert in all the fields underlying the development of an ECA system.

- Representation languages: they model various aspects of information required at different stages of behaviour generation. They aim for the technically detailed annotation for theory-specific information. They are crucial in research contexts, because of the necessity to represent highly specific, low-level information. They ease the integration and exchange of components in multi-modal systems.

- Scripting languages: they incorporate procedural knowledge in their annotations, such as conditional executions of behaviours. They are comparable to high-level programming languages.

- New hybrid formats: in interactive systems, since they need for including at least some procedural capabilities typical for scripting languages into representations.

Even though a considerable overlap between different representations, at the same time they are not directly compatible or cannot be easily adapted for the needs of ECA projects. The need for a common architecture for ECA systems has been faced by SAIBA, an international research initiative whose main aim is to define a standard framework (i.e. a conceptual architecture and associated standard languages) for the generation of virtual agent behaviour (Vilhjálmsson et al., 2007).

In SAIBA initiative, several research groups have joined forces to come up with a commonly agreed on framework for multi-modal behaviour generation. The goal of this initiative was to develop common specifications for representation languages, meant to be:

- Application independent,

- Graphic model independent,

- Different for function and behaviour specification (FML, BML, see below).

SAIBA effort is to create a representational framework for real-time multi-modal behaviour generation in ECAs, by proposing knowledge structures that describe the form and generation of multi-modal communicative behaviour at different levels of abstraction.

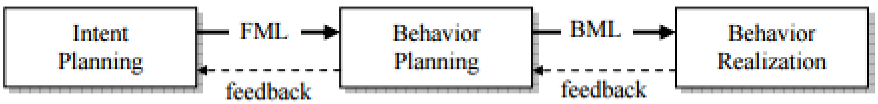


Figure 2.6: The three stages of behaviour generation in the SAIBA framework and the two mediating languages FML and BML. (Vilhjálmsson et al., 2007)

SAIBA framework proposes a pipeline for the generation of ECA behaviour, organizing the different processing steps into three top level modules (Krenn et al., 2011; Vilhjálmsson et al., 2007)

1. INTENT PLANNER (IP): this high-level module concerns agent's communicative intentions, i.e. its goals, emotional states and beliefs, that affect what the agent wants to communicate to the user. It can also include a dialogue planner, which generates an initial abstract version of a dialogue, as a sequence of dialogue acts, without specifying the words, but only the communicative functions of in the dialogue, such as requesting for information, answering

a question, giving feedback, etcâĂę Thus, at this level no reference to any physical or verbal behaviour is specified.

SAIBA framework considers the agent to be a speaker, while some systems, such as GRETA (see 3.1), have implemented intent planners suitable also for listener agents. For example, intent planner could decide that the agent is going to say a sentence, with particular emphasis, being in a certain mood. The main tasks of intent planner are:

- To decide agent's current backchannels (i.e. the communicative intentions);
- To encode such backchannels into Functional Markup Language (FML). This language includes all information regarding agent's mental, affective and affective state that is necessary to create a link between intent and behaviour planning, without reference to surface form. It provides a large spectrum of information including semantic, communicative, discursive, pragmatic and epistemic information.

2. BEHAVIOUR PLANNER (BP): this low-level module conveys the output generated by the previous module (i.e. the agent's intentions), by scheduling a number of communicative signals, such as speech, facial expressions and gestures, encoded with Behaviour Markup Language (BML), an almost standardized language commonly used by ECAs community, that specifies the verbal and non-verbal behaviours, independently of the particular realization (animation) method used. In addition, information about the temporal order and the relative dependencies between the communication channels involved is generated.

   In other words, the Behaviour Planner gives a detailed description of the behaviours (verbal and non-verbal) that should be performed by the agent in order to realize its intentions. These descriptions are written using a specific language, BML, which contains tags specifically aimed to describe behaviours.

   The main subtasks of BP are:

   - Multimodal planning: it accounts for modelling the communication channels and the integration between gestures and syntax on the basis of semantic and pragmatic content of the utterances;
   - Multimodal alignment: it takes care of the timing of the expressions in different channels relative to each other, thus their temporal synchronization, as well as their spatial synchronization.

3. BEHAVIOUR REALIZER (BR): the goal of this last module is to realize the behaviours scheduled in the previous planner by generating animation from the BML inputs. This task involves a series of steps like the synchronization of different modalities and the resolution of conflicts between behaviours of the same modality.

   The main important outputs of BR are:

- Speech generation: thanks to an additional module sounds files are generated given input text annotated with additional information. Pronunciation, prosodic properties, pitch, duration of phonemes can be generated, as well as a temporal information to allow for multimodal synchronization, for instance with visemes (mouth shapes related to sounds/phonemes) and with gestures;
- Gestures realization: gestures are composed starting from basic hand-arm movement trajectories, defined according to the position of the wrist in 3D space, and the hand configuration. More complex gestures can be built by sequences of basic gestures;
- Facial expressions: there exist several coding schemes to describe facial expressions. For example, Facial Animation Parameters (FAPs) are used to represent basic facial actions including tongue, eye and mouth movements, that can combine to represent facial expressions.

In addition, BR gives feedback to the Behaviour Planner about the success or failure of various behaviour requests, in order to successfully plan subsequent actions, especially in real-time systems that incrementally generate behaviour. This important feedback can be of three kinds: success; warning, whenever the BR is unable to 100% comply with a behaviour request; exception, when a behaviour or a block of behaviours had to be cancelled.

2.3 Why laughter in Virtual Agents

At this point it should be fine to open a brief parenthesis about the advantages of building virtual agents that are able to laugh and to interact with laughing users.

Since one the main important functions of interaction design is to reduce learning times, task-performance times and error rates (Morkes et al., 1999; Shneiderman, 1986), off-task activities, such as joking or being playful, were considered for decades diversions from completing the task and a waste of time: “only players of video games and computer games seek entertainment”, Shneiderman wrote, “but applications users focus on their task and may resent too many playful distractions”. Nielsen also advised that, instead of being humorous, tasks assigned in a usability test of productivity software should be business-oriented and as realistic as possible. (Nielsen, 1994).

More recently, another line of research suggested that humour might have positive effects in HCI, as it does in interpersonal interactions and other kinds of communication. This is the point of view of CASA paradigm (see 2.1.1), that argues that users respond to computers in the same ways they respond to other people, because they treat them not as mere tools, but as social actors. Under this perspective, the same positive effects of laughter and humour found in human-human interactions should be elicited also in human-computer interactions. Moreover, some specially experiments designed by Morkes et al. found that participants who received jokes during the interaction with a system rated the system as more likeable and competent, and they responded to it in a more sociable and cooperative manner (Morkes et al., 1999; Nijholt, 2002).

There are many reasons that push towards an integration of laughter in virtual agents. First of all, laughter and humour can be placed in the set of emotions; many authors, like Picard, argue for the integration of emotions into artificial intelligence systems in order to attempt to pass the Turing test. Picard stated "A machine, even limited to text communication, will communicate more effectively with humans if it can perceive and express emotions" (Picard, 1997). The same applies to laughter (we are still considering hilarious laughter): if a virtual agent laughs, this means that it is expressing the emotion of mirth, thus this could improve the quality of its human-like characteristics and naturalness.

Another argument in favour of endowing virtual agents with laughter comes from natural language interfaces: as well as it occurs with among humans, humour could be used to solve communication problems that may arise during the interaction. Humour can help to make the imperfections of natural language interfaces more acceptable for the user and can make them much friendlier.

Finally, according to the theories of humour that support the positive effects of humour and laughter at school and in learning contexts (see 1.2.2), ECAs used as tutors or in e-learning applications should benefit from using humour, as well as user's learning should be improved.

2.4 The Project ILHAIRE

In the first chapter we talked about laughter in humans and its communicational functions; at the beginning of this chapter we presented virtual agents and their increasingly role in interaction with humans and we focused on the potential positive effects of endowing virtual agents with laughter. During last years the idea of integrating laughter in virtual agents and studying their impact in interaction and human experiences was eventually realized thanks to the European project ILHAIRE, i.e. Incorporating Laughter into Human Avatar Interactions: Research and Experiments.

Figure 2.7: The project ILHAIRE's logo.

The ILHAIRE project started in 2011 and lasted 36 months. It has involved researchers from different domains, such as computer scientists, engineers, psychologists and designers, affiliated with several Universities: Research Center in Information Technologies and Numediart Institute for New Media Art Technology at University of Mons (UMONS) - Belgium; LTCI (Laboratoire de Traitement et Communication de l'Information), a joint laboratory between CNRS and TELECOM ParisTech - France; The Laboratory for Human-Centered Multimedia (HCM) at University of

Augsburg - Germany; Casa Paganini - InfoMus at University of Genova - Italy; UCL Interaction Centre of University College London - UK; Emotion research group at Queen's University of Belfast - UK; Personality Psychology and Assessment at the University of Zurich - Switzerland.

As anticipated before, the main purpose of the project was to help the scientific and industrial community to bridge the gap between knowledge on human laughter and its use by agents, thus enabling sociable conversational agents to be designed, using natural-looking and natural-sounding laughter. Data on laughter have been gathered using high quality sound, video, facial and upper body motion capture. The process of database collection has been grounded in psychological foundations and the data has been used to validate computational and theoretical models of analysis and synthesis of audio-visual laughter. Dialogue systems between humans and agents have been developed and studies to capture the qualitative experience evoked by a laughing avatar have been conducted.

2.4.1 Work Packages

The project structure was planned in 7 steps, or Work Packages (WP):

1. WP1: the creation of a database by both using existing resources containing audio-visual records of laughter and generating new ones, by identifying laughter-inducing material in order to design experiments to record natural laughter to add to the existing resources. The goal of this initial phase was to build an annotated database of multi-modal records of laughter in naturalistic interactions, in order to label different types and functions of laughter.

2. WP2: a multi-modal analysis of hilarious and conversational (social) laughter (see 1.1.1), in order to investigate the acoustic features, facial actions, expressive gesture features, influence of culture and gender. The goal of his step was to enable automatic detection and classification of laughter.

3. WP3: the generation and synthesis of audio-visual laughter, relying on the annotated database built in WP1 and the analyses of WP2.

4. WP4: the design of an adaptive data-driven multi-modal dialogue management system for achieving natural and user-friendly interactions integrating laughter. During this phase Machine Learning techniques were used, considering WP1 data as baseline, and input from WP2 for adapting the strategy online afterwards. In addition information to the laughter synthesis system developed in WP3 was provided.

5. WP5: the psychological foundations of laughter, with the goal of understanding the factors affecting the conveyance and perception of laughter expressions, and developing psychologically appropriate methods (quantitative and qualitative) of assessing affective and cognitive responses to laughing agents.

6. WP6: the integration of the various components developed within the project, and the evaluation of the integrated systems; both quantitative and qualitative approaches (e.g., semi-structured interviews, effect on task performance) were

used to evaluate the level of emotional contagion triggered by the agent's laughter expressions.

7. WP7: the creation of the web site for the dissemination of project results [1].

2.5 Laughter detection and response systems

It is time to provide some more details about the recent attempts to build virtual agents capable of laughing. Most of the projects that are going to be presented below were developed in the context of the project ILHAIRE.

2.5.1 AVLaughterCycle

Starting in 2010, a large group of researchers was involved in the AudioVisualLaughterCycle project, with the purpose to create "an audio-visual laughing machine capable of recording user's laughter and to respond to it with a machine-generated laughter linked with the input one" (Urbain et al., 2010), in order to create an engaging interaction loop between human and agent, aimed to transform the initial forced laughter of the user into spontaneous one.

Their work was the first attempt to provide both audio recording and facial motion tracking of laughter, while the previous works had focused only on audio features. To train their systems, a database was first created, called AVLC corpus: in this audio-visual database of human laughter, in order to elicit laughter as spontaneous and natural as possible, they chose to display funny movies to the users. When recording laughter in front of a camera, in fact, it is hard for participants to laugh spontaneously, either because of the presence of markers in the face or the body due to motion capture technique, or because of the limitations due to the camera space, or at least because they know to be recorded. In other cases it is fine to take laughter recordings while collecting data for other purposes, for instance in corpora of speech. Such cases are not the best for analysis of hilarious laughter, because they are more probable to concern social and conversational laughter.

Basically, the system was built to detect laughter, offline from an audio file, or in real-time, and, after analysing its features, to choose from the database the laughter most similar to the input.

To summarize, the architecture of AVLaughterCycle includes :

- Smart Sensor Integration (SSI): a software to collect, synchronize and process from different data sources. Given the audio laughter segment as input, SSI processes it by extracting features vectors, containing both spectral and temporal features.

- MediaCycle: this software receives the features vector representing the input audio file, and measures the similarity (using Euclidean distance) with annotated laughs of AVLC database. The comparison with the database allows for the selection of the output, that is the most similar laughter.

[1] http://www.ilhaire.eu/

- Motion capture. The database is annotated also for facial features: in this way, when the output laughter is selected, it is possible to use the facial data corresponding to that audio. They will be then mapped (using linear and weighted functions) to standards parameters and adapted to the face of the virtual agent.

This first attempt presented some limits, for instance it selected the output according to similarity criteria, even though it is not probably the natural way of joining somebody laughter. In addition, the laughter was not synthesized since it was selected from the database.

2.5.2 Laugh Machine

In 2012, starting from eNTERFACE'12 Workshop, a new interactive system was developed, called Laugh Machine: a complete interactive system enabled to detect human laughter and respond appropriately, with right timing and intensity, by integrating information of human behaviour and the context (Urbain et al., 2013a). Laugh Machine exploited the AVLC corpus built during the previous project, but it presented some improvements with respect to it, such as the introduction of laughter intensity and context information as elements which modulate the output response.

The architecture of the system is more complex than the previous one; the different modules are organized in three phases:

1. Input: this step includes multi-modal data acquisition and real-time analysis using SSI. The input data are :
    - User's laughter detection,
    - Laughter intensity estimation exploiting previous manually annotations,
    - Context information, about the funniness of the content to which the user is exposed.
2. Decision making: this is the main step of the process, aimed at controlling the agent audio-visual response, by deciding when and how it laughs in order to generate a natural response able to elicit positive humour-related affect in the user, and emotional contagion. The decision making process consists of two modules:
    - Dialogue Manager: it concerns decisions independent of the agent's characteristics. At each time frame, it receives as input the information about intensity and likelihood of laughter, along with the contextual ones, in form of vector of features; then, following a decision rule, it generates as output a two-dimensional vector, containing the laughter intensity and duration. The laughter response is chosen from an output cluster, that is a set of possible vectors generated by the decision rule: in this way it is possible to maintain some variability;
    - Laughter planner: this module follows the dialogue manager, and it differs from it since it depends on the characteristics of the agent that will display the output (e.g. the gender). Laughter planner controls

the details of the expressive pattern of the output laughter response, by selecting appropriate audio-visual episodes from the pre-synthesized laughter samples, and encoding them in the appropriate language. In other words, this module adapts the outputs generated by dialogue manager to the constraints of synthesis modules, by choosing the example which best matches (in term of less Euclidean distance) the values of intensity and duration, and context information.

3. Output: it is the visualization of the audio-visual laughter synthesis that best integrates the information from context and user laughter.

2.5.3 Multimodal expression of laughter

More and more efforts are being made towards improving the synchronization between different modalities of expression, such as audio and visual ones. In this context, Niewiadomski and Pelachaud have given a strong contribution during their studies, where they focused on modelling laughter with different intensities, and they focused also on respiration behaviour synthesis, never considered before in virtual agents (R. Niewiadomski & Pelachaud, 2012). Even though the concept of intensity was already considered in the previous work (2.5.2), the approach of the authors was novel than the previous one. For intensity, they exploited human annotation of the perceived intensity of the laughter episodes in AVLC corpus, using Likert scales. From these data, they investigated the correlation between facial cues (in terms of distances between facial markers corresponding to specific muscular activation) and the perception of intensity level. The highest correlation was found in the action units (see 1.3.1) corresponding to mouth opening, i.e. the higher the intensity is, the more open is the mouth.

Rather than a simple linear increase of the activation of all facial cues, intense laughter was found to be expressed by activating some additional AUs or by gradual changes of single AUs. In other words, only some AUs change and their modifications are independent to each other. By taking into account these findings, the authors proposed a novel approach for real-time intensity modulation, where there is not a simple linear filter over the whole face, but facial parts are differently modulated according to empirical results (R. Niewiadomski & Pelachaud, 2012).

Regarding respiration modelling, they studied the relation between facial and audio patterns and respiration phases (inhalation and exhalation), in order to synchronize them. For example, mouth opening is more intense in exhalation, because it allows for air expulsion, as well as the backward tilt of the head facilitates the forced exhalation. They used manual annotation of respiration phases and airflow direction, while in following works other devices will be used, such as a respiration sensor to measure thoracic circumference (Urbain et al., 2013b). Since the standards of animation for virtual agents only allow to animate the body skeleton, but not the body shape, to simulate respiration phases the authors added two bones in the thorax.

2.5.4 Laughing when you are winning

In 2013, during eNTERFACE Workshop, a new project, built upon the Laugh machine, introduced important innovations (Mancini et al., 2014).

First, laughter detection was expanded to multi-modal decision, analysing audio, facial and body movements. The laughter recognition component SSI was used to synchronize the data acquisition from different sensors and computers, to estimate users' states coming from different modalities (audio, face analysis and body analysis), and to fuse them together. Thanks to SSI, raw data were filtered from noise and then features extraction was performed; the resulting data could be classified using thresholds or learning, and a deeper analysis was possible when taking into consideration other elements, such as past events or events from other modalities. In this way a meaningful interpretation of the detected events was possible. During the fusion of the data, new vectors were generated with every new event: only if several successive events were pointing in the same direction, the fusion vector would be dragged into that direction. Hence, the robustness of the detection system was highly improved thanks to the new fusion techniques.

With respect to face analysis, the probability of smile was computed by combining the most promising AUs for smile (upper lip corner, lip stretcher and the negative of lip corner depressor) to a single value; a happy score was also derived from the detected face, since it has been found to be highly correlated with smile.

Regarding body analysis, the authors improved and extended previous techniques by introducing a new information about laughter intensity. Mancini et al. computed the Body Laughter Index, a weighted sum of the correlation of vertical positions of the shoulders and the kinetic energy computed from the speed of the shoulders and their mass relatively to body mass (Mancini et al., 2012). In "Laughing when you are winning", this model was improved by an index approximating laughter intensity. This index considered periodic movements and non-periodic ones of shoulders and trunk, and significant correlations were found between this index and the manually annotated intensity values.

Another important novelty of the system concerns the integration of laughter mimicry. A mimicry module was built with the function to decide how the agent's expressive behaviour should mimic the user's one, that is, modulate its quality of movement (amplitude, speed, fluidity, energy, etc.) depending on a set of expressivity parameters. The module receives information about user's body movement features, laughter probability and laughter intensity, resulting from the fusion process.

Part II

Building and evaluating a laughing Virtual Agent with copying behaviour.

Chapter 3

The Laughing Agent

A truly interactive machine should show not only synchrony with the user, that is, a behavioural adaptation toward a shared behaviour, but also copying the user behaviour and its expressivity (Broekema, 2011). These two types of coordination characterize human-human interaction and should be replicated to guarantee an effective human-machine interaction. In fact, as argued by Nagaoka et al.(Nagaoka, Komori, & Yoshikawa, 2007), nonverbal synchrony is a signal of the user's attitude toward other users or the interaction, while Chartrand and Bargh (Chartrand & Bargh, 1994) affirmed that copying behaviour is exhibited to gain social influence. Bailenson and Yee (J. Bailenson & Yee, 2005) confirmed such an effect (i.e., to gain social influence) in a study where a virtual agent copied head movement of a human confederate. Broekema (Broekema, 2011) investigated on the effects of intensity copying of head movements in virtual agents. Results showed that such copying has a positive influence on the perceived image of the avatar. Buschmeier et al. (Buschmeier, Bergmann, & Kopp, 2010) proposed an agent gesture and vocal model for syntactic, lexical and gestural alignment with the user. Other works, focused more on expressivity copying: Castellano et al. investigated the physical characteristics of human motion (e.g., velocity, acceleration, fluidity) that contribute to convey affective content. To do that, they implemented a virtual agent that performed copying behaviour of actors expressing basic emotional states with voice and gestures (Castellano et al., 2012). It is worth to notice that the agent did not copy the particular gesture performed by the actor, but only the *quality* of movement. Results showed that different movement characteristics can communicate different emotional states, for example anger can be effectively communicated by increasing speed and acceleration of movements.

In our work, we extended the paradigm of (Castellano et al., 2012) to allow for copying expressivity on the fly, that is copying the expressivity dynamically as it evolves in human's performance. Our aim was to study how a virtual agent able to copy and to adapt its laughing and expressive behaviours on the fly participates in enhancing user's experience in the interaction. In the experiment, we tried to make the participant laugh using an audio file and a virtual agent. The agent laughed several times with the music, encouraging the participant to do so. In one condition, the quality of the agent's behaviour was driven by the participant's behaviour quality, mimicking the way he/she laughed.

We used the ECA Greta as the virtual companion of the user, and a specific architecture of the system was implemented in order to endow the agent with laughter and copying capabilities. In the following paragraphs we will describe the standard structure of Greta in 3.1, then we will briefly illustrate the laughter synthesis applied to Greta in 3.2 and finally the architecture of the system used for our experiment is described in 3.3.

3.1 Greta

GRETA is an ECA developed from 2005 by Catherine Pelachaud and her team at Télécom Paris-Tech (Poggi et al., 2005; Mancini et al., 2008; Radoslaw Niewiadomski et al., 2009). It is an implementation of SAIBA framework (see 2.2), with some differences and additions related to the specific goals of the system.

The following description about Greta's architecture and its module mainly relies on the works of (Mancini et al., 2008; Bevacqua et al., 2010, 2007). Greta's architecture takes into account some relevant issues about interaction and tries to implement them:

- First of all, during a conversation an agent may not only speak to the user, but it can have also the role of a listener. In order to allow Greta to be a listener agent, the Intent Planner module, initially thought only for speaker agents, has been split into two elements: the Speaker Intent Planner (SIP), that at the moment is only pre-defined manually, and the Listener Intent Planner (LIP), whose goal is to compute the backchannels of the agent while listening to the user.
- Another important observation concerns the fact that people have different characters and personalities, thus they can express the same attitude or behaviour in different ways and intensities. In order to be aware of this different expressivity when building the behaviours of the ECA, a second source of information has been created to give information to the Behaviour Planner: it is called Baseline, and gives information about the general tendency of the agent to use modalities and expressivity.
- The modules are not independent, instead they can communicate through Psyclone [1], a messaging platform for building complex automation and autonomous systems, and they are synchronized by a Central Clock.

The main elements of Greta's architecture are going to be described into more details, as well as the types of languages used to encode the information and pass it from module to module. The scheme of the architecture is showed in Figure 3.1.

3.1.1 Listener Intention Planner (LIP)

As stated before, the LIP produces as output the communicative intentions of the agent while listening to the user. The output is described in FML-APML language, an XML-based markup language which encodes communicative and emotional functions that the agent aims to transmit, as well as the text to be uttered. The output is

[1] http://www.cmlabs.com/psyclone/

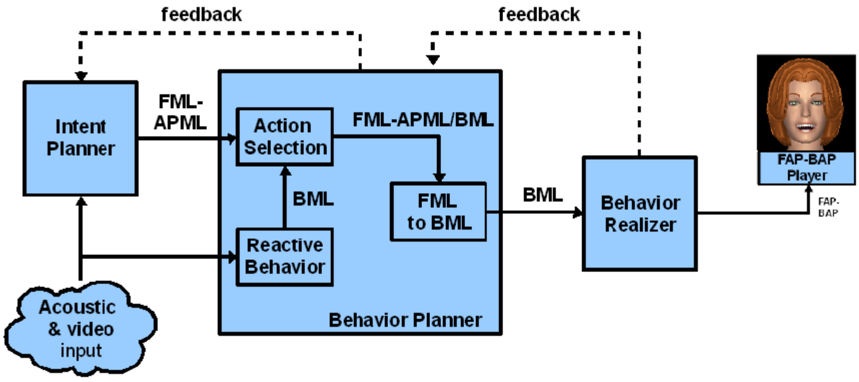


Figure 3.1: Greta's architecture (Mancini et al., 2008; Bevacqua et al., 2010)

generated through algorithms following probabilistic rules, applied to three inputs: user's verbal and non-verbal behaviours (captured by video camera and microphone), estimated interest level (obtained through gaze, head and torso data) and mental states. There exist three types of backchannels, according to the temporal line:

- Reactive backchannels: they come from the first process of the perception of user's speech;
- Response backchannels: the come from a more aware evaluation of the inputs, and thus they comprehend memory and cognitive processes;
- Mimicry backchannels: they come from imitation of speaker's behaviour.

3.1.2 Behaviour Planner (BP)

It takes as input both the agent's communicative intentions specified by the FML-APML language and some agent's characteristics (i.e. baseline), and produces an output described in the BML language (see 2.2). It selects for each communicative intention the adequate set of behaviours to display, as well as their timing information.

The interaction between an ECA and a human requires a mutual adaptation, such as reactivity to speech, changes of facial expressions, reactions to the situation and adaptation to user's behaviours. To implement this dynamical part of the social agent, the ECA needs reactivity (real-time perception) and sensitivity (real-time adapted actions). An architecture enabling dynamical coupling has been adapted to Greta independently from SAIBA framework. This architecture implements two capacities giving to the agent dynamical properties and reactivity: the self-generation of dynamics (driven by an oscillator) and the sensibility to partner's behaviour (detected by movement analysis). When two agents interact with each other, coupling behaviours such as synchronisation and turn-taking emerge (Bevacqua et al., 2010). The authors proposed the implementation of two important modules:

- *Reactive Behaviour* module: it is a complementary part of the Intent Planner, much more reactive and also working at a much lower level; it acts more as an adaptor of the ongoing interaction than as a planner. It is relatively autonomous from the rest of the architecture, since it receives inputs directly from audio-visual data, without passing through the IP; it receives as well the currently planned actions output by the Behaviour Planner. With these two sources of information, the Reactive Behaviour module will propose to the Action Selection module (see below) two different types of data: it can propose adaptation of the current behaviour by comparing its own actions to the actions of the speaker at a very low level, and, by extracting from the user's behaviour salient events such as facial expressions, speech prosody or breaks, it can propose actions such as performing a backchannel, imitating the user, smiling etc. Moreover, the Reactive Behaviour can propose real-time reactions or adaptations to the user's behaviour thanks to its partial autonomy.

- *Action Selection* module: it selects the action to display, by merging the information coming from both the Reactive Behaviour and the Intent Planner. Thus, it has to choose between a more cognitive-driven and a more reactive-driven behaviour.

3.1.3 Behaviour Realizer (BR)

It receives the BML input about the text to be spoken and/or a set of non-verbal signals to be displayed, and generates the animation of the agent, following MPEG-4 format. The main tasks of BR are:

- Generation of the speech by the Text-to-Speech module;
- Lips movements generations;
- Synchronization between different modalities;
- Generation of Facial Action Parameters and Behaviour Action Parameters frames.

Finally, it sends the animation data to FAP-BAP Player, that plays them in a graphic window.

In addition, Greta is expressive ECA, thus its animation can be qualitatively modified by a set of expressivity parameters affecting the physical characteristics of movements (like speed, width, strength, etc...). (Bevacqua et al., 2007) defined Greta's expressivity by six dimensions:

- Overall Activity: it models the general amount of activity (e.g., passive/static or animated/engaged);
- Spatial Extent: it modifies the amplitude of movements (e.g., expanded versus contracted);
- Temporal Extent: it changes the duration of movements (e.g., quick versus sustained actions);
- Fluidity: it influences the smoothness and continuity of movement (e.g., smooth, graceful versus sudden, jerky);

- Power: it represents the dynamic properties of the movement (e.g., weak/relaxed versus strong/tense);

- Repetitivity: it models the tendency of the agent to replicate the same movement with short and close repetitions during time.

To summarize, Greta is an ECA:

- Endowed with expressive capabilities,

- Communicating its mental and emotional states using different modalities,

- With own behaviour tendency,

- Able to perform mimicry, synchronization and engagement with the user.

In our experiment, we personalized the role of Greta from a virtual conversational companion to a virtual companion engaged in a laughter experience; the laughter synthesis and the architecture of the system developed for this purpose will be described in the next paragraphs.

3.2 Laughter Animation Model

During laughter the whole body moves. The synthesized animation model used for Greta involved all upper body modalities (except gaze): facial expression (eyebrow, eyelid, cheek, lip and jaw), head and upper torso motions. It relied on previous data-driven works (Ding et al., 2014). The laughter animation generator took as input laughter audio signals, such as its transcription into pseudo-phoneme (in reference to speech phoneme) including its duration and its intensity, as well as two speech prosodic features, namely pitch and energy. Three different models to generate laughter animations were derived: the lip and jaw animation synthesis; upper facial expression synthesis; head and torso animation synthesis. We had to define 3 different models as we did not have a single database that included both facial data and body data. We used two databases, one for facial and lip motions and one for upper body motions.

3.2.1 Lip and Jaw Animation

During the learning phase, linear regression models were used to design the lip and jaw motion synthesis system (more details can be found in (Ding et al., 2014)). The underlying idea of a linear regression model is to estimate a desired quantity x (the lip and jaw motion) as a linear function of an observed quantity θ (the speech features).

Such linear regression models were used for each of the 14 laughter pseudo-phonemes so that a set of 14 linear regression models was built.

Each linear regression model was learned to model the dependencies between the lip/jaw motion and the speech features from a collection of training pairs of speech features and of lip and jaw motions. The method of least squares was used to learn the linear regression model of each laughter phoneme.

During the synthesis phase, the animator generator module took as inputs a sequence of prosodic features and of laughter pseudo-phonemes and their duration. The synthesis of the lip and jaw motion was computed independently for every laughter phoneme of the sequence. Then, the obtained animation for the lip shape was smoothed between successive laughter pseudo-phonemes to simulate co-articulation effect.

3.2.2 Upper Facial Expression

The upper facial expression synthesis was based on the concatenation of human segmented motions. The model described in (Ding et al., 2014) was extended to synthesize the motion of upper facial expressions (eyebrows, eyelids and cheeks). The synthesis by concatenation approach consists in selecting and concatenating motions from human motion data corresponding to the input laughter pseudo-phonemes sequence. That is, for each of the 14 laughter pseudo-phoneme labels, pp_i, a number N_i of human upper face expression, called $S_i = \left\{m^i_j, j = 1..N_i\right\}$, can be extracted and collected from the dataset of motion capture of human laughter.

The algorithm worked as follow. Given an input sequence of laughter pseudo-phonemes of length K, $(p_1, ...p_K)$ (with $\forall k \in 1..K, p_k \in \{pp_1, ..., pp_{14}\}$) and $d(p_k)$ the duration of the k^{th} laughter pseudo-phoneme in the sequence, the *synthesis by concatenation* method found the best sequence of segments $(s_1, s_2, ..., s_K)$ belonging to $S_{p_1} \times S_{p_2} \times ... \times S_{p_K}$ (with $d(s_k)$ the duration of the segment s_k). For this computation it minimized a cost function that was defined as the weighted sum of two factors: duration cost and continuity cost. The former increased with the difference between the lengths of a human segmented motion and of the corresponding laughter pseudo-phoneme; the latter increased with the distance between the last position of a selected segment and the first position of the succeeding segment. The cost function represented the quality of fit between the segment sequence and the laughter pseudo-phonemes sequence.

3.2.3 Head and Trunk Animation

Head and torso motions are characterized by 3D rotation angles (hence a 6 dimensional signal). The synthesized animation model took as input a sequence of laughter pseudo-phonemes together with their duration and their intensity (low or high), and prosodic features (pitch and energy). It output values of the 3D rotations for the head and torso. The model relied on the method of (Ding et al., 2014), where an extension of Hidden Markov Model (HMM), called Coupled Transition Parameterized Loop HMMs, was proposed as animation generators. This method allowed to generate 3 types of motions.

- Modelling shaking-like movement (for shoulder and torso movements),
- Emphasizing prosodic features influence on the synthesized animation,
- Taking into account the dependencies between the movements across modalities (head and torso).

The synthesis was performed through two steps. First a state sequence was determined according to the transition probability distribution (a state represented a clustering of one-dimensional motion trajectory and was modelled by Gaussian distribution). Then, an animation sequence was estimated as the most likely observation sequence given the state sequence; it consisted in the sequence of the means of the Gaussian distribution of the states in the sequence.

3.3 General Architecture

The architecture of the laughing agent is illustrated in Figure 3.2. Two different platforms were used: the *Detection Platform*, implemented via EyesWeb XMI platform [2], a modular system that allowed both expert (e.g. researchers in computer engineering) and non-expert users (e.g. artists) to create multi-modal application in a visual way; the second platform was the *Virtual Agent*, designed using the Greta agent platform described in 3.1. The two platforms communicated via ActiveMQ [3] messages.

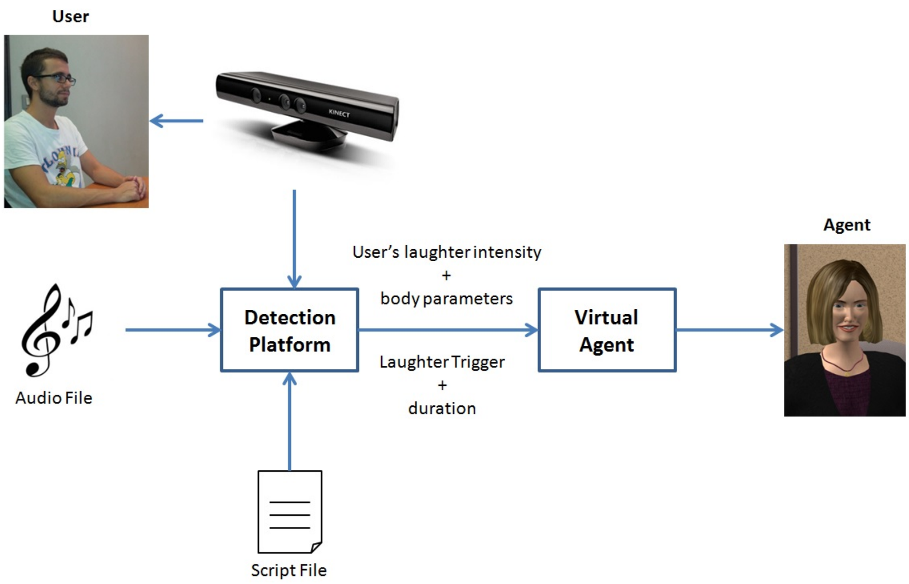


Figure 3.2: The architecture of our laughing agent.

3.3.1 Detection Platform

EyesWeb is conceived, designed and developed by InfoMus Lab of Genova. It supports a wide number of input devices including motion capture systems, various types of professional and low cost video cameras, game interfaces (e.g., Kinect, Wii), multichannel audio input (e.g. microphones), analog inputs (e.g. for physiological signals). Supported outputs include multichannel audio, video, analog devices,

[2] http://www.infomus.org/eyesweb_ita.php
[3] http://activemq.apache.org/

robotic platforms. EyesWeb supports real-time synchronized recordings of multimodal channels, and includes a number of software libraries, including the Non-Verbal Expressive Gesture Analysis and the Non-Verbal Social Signals Analysis libs.

The detection platform received as inputs the information about the user, an audio file and a script file. The script file was designed since it was not possible for us to automatically compute which features of the music stimuli should trigger a laughter, and how funny it was supposed to be, thus we needed a file (one per audio file) containing three important informations:

1. time markers to know when to trigger laughter,
2. duration of laughter,
3. intensity of laughter.

Such markers were needed as the architecture did not include a semantic analyser of audio, so we were not able to automatically detect when the audio stimulus was funny enough to trigger laughter.

The platform had two purposes for the experiment. First, to play an audio file, and to send triggers for laughter at predetermined times: every time one time marker was reached, one message containing the duration and the intensity of the laughter was sent to the Greta.

The second purpose of the detection platform was to detect the user's body features and thus to compute the user's laughter intensity. These information were continuously transmitted to our virtual agent to make it able to copy the user's behaviour quality.

Features extraction was computed from the BW depth map video (each pixel was a 16 bit value indicating the distance from the camera) captured by a Kinect [4] sensor (640x480 @ 30 fps). The Microfost Kinect SDK extracted the user's silhouette from the background image and provided it as output. Head's and shoulder's positions were determined by recursively segmenting the user's silhouette in EyesWeb. Figure 3.3 shows the results of this process, where H, LS and RS are the head, left shoulder, and right shoulder position, respectively.

For the experiment, we took two parameters into account to drive the agent's behaviour:

1. User's body leaning, directly mapped to the agent's body leaning:
    - Front/back leaning was estimated by computing the mean color of the depth image of the user provided by Kinect. In the image, each pixel had a value in the range $[0, 65535]$ depending on its distance from the sensor. We split the user's silhouette in two halves separated by the horizontal line passing through the user's barycentre (see Figure 3.4 in which we highlighted the pixels belonging to the user's upper and lower body halves). For both halves, we computed the mean color of the pixels belonging to each area. Finally, the front/back leaning was the ratio between these 2 values.

[4] http://www.microsoft.com/en-us/kinectforwindows

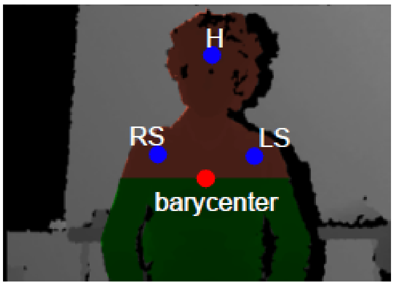


Figure 3.3: Input RGB and depth images were preprocessed to isolate user's body upper/lower part (red and green areas) and head/shoulder centroids (blue dots).

Figure 3.4: Computation of front/back leaning as the ration between the mean color of the upper and lower body halves.

- Left/Right leaning was approximated by the left/right symmetry of user's body silhouette. Body symmetry is a relevant feature for detecting emotional states: facial expressions exhibit right/left asymmetries that are emotion-dependent (Borod et al., 1997); emotional states influence the asymmetry of body movements (Roether, Omlor, & Giese, 2008); arm position symmetry reveals "relax" attitude of a person within a group (Mehrabian, 2007). Figure 3.5 shows how we computed left/right symmetry. The vertical projection H' of the head position H on the segment joining right and left shoulder $\overline{RS, LS}$ was computed, the triangle $H, \widehat{H'}, LS$ was mirrored with respect to the vertical axis and then subtracted from the triangle $H, \widehat{H'}, RS$. The area of the resulting shape was normalized in the range $[0, 1]$ by dividing it by the sum of the area of the 2 triangles.

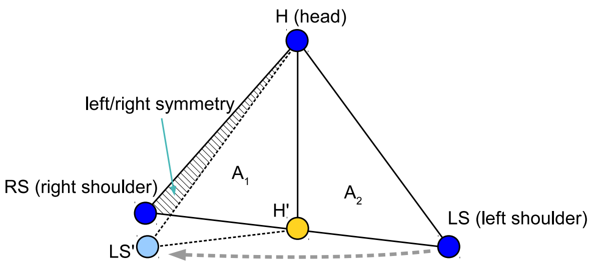


Figure 3.5: Computation of left/right symmetry as area of the group of pixels resulting from the subtraction between the triangle $H, \widehat{H'}, RS$ and the triangle $H, \widehat{H'}, LS'$.

2. User's laughter intensity: while there was a direct mapping between the user's and the agent's leaning, the user's laughing intensity had an overall influence on the agent's body movements.

   As explained above, the user's body leaning and his laughing intensity influ-

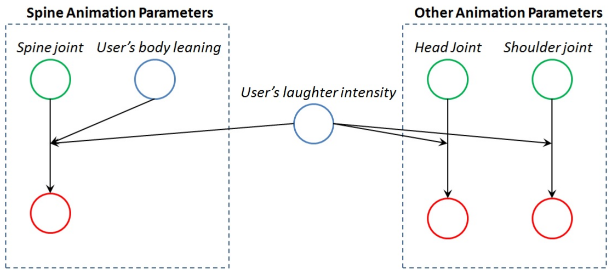


Figure 3.6: In green, animation parameters generated from our model. In blue, user's parameters. In red, final animation parameters.

enced in real time the animation of the agent. If the user did not laugh at all and/or did not move at all, the agent's behaviour would be inhibited, and the agent would stay still. On the other hand, a user laughing out loud would lead the agent to move even more. In the *yes_Copying* condition, that is where the agent was able to adapt its laughter according to the user's behaviour (see 4.4), the user's body parameters modified the agent's animation.

To modify the generated animation of the agent on the fly, a graphical tool was designed, allowing for the manipulation of different inputs and for blending them to obtain a new animation as output (see Figure 3.6). Here, two types of inputs were blended:

1. Animation parameters generated by the laughter animation model
2. User's parameters (body leaning and laughter intensity)

These inputs are respectively represented in Figure 3.6 by green circles and blue circles. The red circles represent the final animation parameters played by the agent.

In the *no_Copying* configuration where the agent did not adapt its behaviour (see 4.4.1.2), the generated animation parameters from the synthesis model were simply transmitted to and displayed by the agent. In the *yes_Copying* configuration, however, these animations parameters were influenced by the user's parameters. User's body leaning influenced the spine animation parameters: if the user leaned forward while laughing, then the agent did the same. User's laughter intensity influenced not only the spine animation parameters, but also all the other body animation parameters generated by the laughter synthesis model: if the user laughed with large movements, the amplitude of the agent's movements was increased. On the other hand, if the user did not laugh at all or performed small motions, then the amplitude of the agent's laughter motions was decreased; it could decrease up to extremely low amplitude making the agent appear to stay still.

Chapter 4

The Evaluation Study

The aim of the evaluation study was to evaluate the performance of the system, and to investigate whether and how the presence of a laughing virtual agent, either copying or not user's laughter intensity, affects user's experience while listening to a funny music. In order to better analyse this impact, we considered three main hypotheses, involving the variables we assume to affect the experience with the agent, such as presence and behaviour of the agent, mood and personality of the user (see 4.5. The originality of our study relies to the fact that, differently from previous studies aimed to assess the influence of a virtual agent on user's amusement, where laughter the typology of stimuli used to elicit laughter in the user were videos, comics or verbal jokes and puns, only audio stimuli were created. Moreover, the choice of using an audio stimuli as laughter-eliciting signals was made in order to eliminate any visual factors that could influence and distract participants' behaviours: with audio-only stimuli, in fact, participants might spend more time looking at the agent, and, as consequence, might be more attentive to its expressive nuances.

In the next paragraphs the experiment will be described into details; a brief description of the scenario presented to the user is provided in 4.1, the set-up of the experiment room is described in 4.2, followed by the experimental protocol in 4.3 and the design in 4.4. Finally, data analysis and its results will be showed in 4.6 and 4.7, and discussed in 4.8. I was involved in most of the phases of the study, except for the preparation of the stimuli, built by InfoMus group of University Genova, and the management of the general architecture (Greta, Eyesweb, ActiveMQ), for which I was helped by a PhD student of our group.

4.1 Scenario

Like the majority of experiments, our study was carried out in laboratory environment. One of the main problems of this kind of settings is the difficulty for the participant to behave in a natural way, like he/she was in a familiar place. In our context, it could be hard to spontaneous laugh when being in an unfamiliar place, being aware to be registered. The participant could for example be afraid to show his/her amusement, since he/she could think to be inappropriate or to disturb the process of the experiment. Since our main purpose was to make the users laugh to measure their laughter intensity, we needed a way to encourage them to laugh, at the same time without reveal what we would measure.

Thus, we prepared a cover story (see 1.3.3) to tell the participants during the instructions of the experiment. First of all, Greta was introduced as a synthetic computer animated character designed to autonomously interact with humans in various activities. Although most of the participants were used to virtual agents, we provided the instructions with a brief description of them, for those users who did not know about them. Participants were instructed to consider Greta as a colleague sitting on a desk in the same room: in this way the problem of unfamiliar and strange room could partially be reduced, and the user can imagine to stay in a familiar place with a person who he/she knows. The "fake" research goal which the participants focused on consisted in the task of improving the agent's non-verbal behaviour; in particular, its ability of exploiting user's laughter to improve its interaction with humans. The users were thus asked to teach Greta how and when to laugh, in a natural, effective and convincing way.

4.2 Set-up

As experiments standards, the experiment room and its equipment were the same for each participant. The set-up aimed to recall an open space office, where the desk of the participant and of Greta were set side-by-side. The room equipped with:

- a desk, the location of the user during all the experiment;
- a laptop, on the desk, used by the user to fill in questionnaires and go on the various trials of the experiment;
- two loudspeakers, at the left and the right of the laptop, producing both the audio files and the voice of Greta;
- one big vertical screen at the left of the desk, displaying the virtual agent when present, otherwise being black with the instruction to look at the laptop;
- another computer, not accessible by the participant, containing Greta's commands and Eyesweb; during the experiment, this computer was monitored by the experimenter from another room, thanks to VSNviewer;
- a Kinect sensor, placed in front of the user, used to capture the BW depth image of the participant (see 3.3.1).

During the whole experiment, the participants had to focus only on the laptop and the big screen, according to the instructions present one of them; the user was trained about how to interact with the laptop and the big screen thanks a short tutorial that explained him/her how to switch between each other and to learn which task was associated to each screen: the task of filling in the questionnaires while using the laptop, and the task to teach Greta how to laugh while looking at the big screen.

4.3 Experimental Protocol

The experiment lasted about 25 minutes (the complete schedule of the protocol can be found in Figure 4.1). Participants listened to the same piece of music twice: first alone, and then with the presence of their virtual colleague Greta. At the end of each listening, some questionnaires were proposed to the participants (see 4.3.2), each one for measuring one specific Dependent Variable (DV, see 4.4.2). Finally, some information about participants' personality traits was collected, as well as demographics information.

After the instructions and the tutorial, the user was asked to confirm his/her willing to continue the experiment by signing the consent form, which asked permission to video record the study, use it for research purposes, always letting all personal information be anonymous.

The experimenter stayed in the room only during the instructions about the experiment and during the tutorial, then he went in another room, from which he could control both the computer and the laptop using UltraVNC [1]. By controlling the former, the experimenter could ensure the correct progress of the experiment, changing what was present in the big screen, according to the different steps, while the laptop was monitored only to know when the subject finished to fill in the questionnaires, which was a necessary information to pass from one step to the following one.

[1] http://www.uvnc.com/

Protocol:	Big screen	Laptop	Commands on computer	Duration	Task of the user
Consensus form and Brief Instructions explaining the Context for the trial	Black screen	Instructions		**[3min]**	Listen
Quick tutorial to make the user familiar with the action of switching from one screen to another and learn which task is associated to each screen	Funny video	Black screen	Launch the video	**[5min]**	Watch video Click when finished
	Black screen	Questionnaire	Close the video		Fill in quest- Click when finished
CALIBRATION:	Black screen	Black screen		**[2min]**	Calibration Click when finished
Questionnaire about mood	Black screen	q_mood	Control when he finishes	**[1min]**	Fill in quest. submit
Listening to the music without Greta	Black screen	Black screen	Launch Greta and EyesWeb , minimizing Greta	**[2min]**	Listen to the music
Questionnaire about audio	Black screen	q_audio		**[1min]**	Fill in quest submit
Questionnaire about mood	Black screen	q_mood	Control when he finishes	**[1min]**	Fill in quest submit
Listening to the music with Greta	Greta	Black screen	Launch EyesWeb, not minimizing Greta	**[2min]**	Listen to the music
Questionnaire about audio	Black screen	q_audio	Close Greta	**[1min]**	Fill in quest submit
Questionnaire about mood	Black screen	q_mood		**[1min]**	Fill in quest submit
Questionnaire about Greta	Black screen	q_greta		**[2min]**	Fill in quest submit
Questionnaire about personality	Black screen	q_perso		**[5min]**	Fill in quest submit
Demographics (age, gender, nationality, education, music abilities)	Black screen	demographics		**[1min]**	Fill in quest submit
Debriefing	Black screen	Thank you and debriefing		**[1min]**	Listen

Figure 4.1: The experimental protocol. The table shows what happened during each step of the study, what was displayed on the big screen and on the laptop, and the tasks of both users and experimenter, along with an estimate of the timings. The steps on grey background needed the presence of the experimenter in the room.

As soon as the participant entered the room, the experimenter introduced the experiment, reading a prefixed text also available on the laptop's screen, about virtual agents and the scenario. The big screen's background was black. After the explanation, the participant decided whether to start the experiment by signing the consent form. This preliminary step lasted about 3 minutes. The following 5 minutes were dedicated to the tutorial: the participant watched a funny video on the big screen, while the laptop showed a black screen with the instruction of looking at the big screen. When the video finished, the participant, by clicking a button, approached for the first time to an example of questionnaire, which he was helped to fill in. The tutorial was important because the participant could understand the roles of the big screen and the laptop: in the former he would "see things", while the laptop would be the device used to answer to the questionnaires and to switch from one step to another. When the participant was ready, the experimenter could perform his last action in the room: the calibration of the Kinect with EyesWeb. This step required about 2 minutes, during which both the big screen and the laptop showed black screen, and the subject followed the commands of the experimenter. The real experiment started when the experimenter left the room, and the subject was asked to fill in the first questionnaire about current mood. This 1-minute task involved only the laptop, while the big screen continued to be black. When the questionnaire was finished, the subject clicked a button and the laptop displayed a black screen describing the next step. The experimenter constantly monitored what was happening on the laptop, and when he saw that the questionnaire had been filled in by the subject, he could launch the first condition of the experiment, minimizing the window with Greta, because the first listening required the subject to be alone. The music lasted about 2 minutes, after which the subject could fill in the questionnaires about the funniness of the music and his mood during the experience. Each questionnaire lasted about 1 minute, during which the black screen was still black. The next step required again the action of the experimenter, who launched the second condition, in this case expanding the window of Greta on the big screen, while the laptop displayed the instruction to look at the big screen. After 2 minutes of listening, Greta was closed, and the big screen remained black for the rest of the experiment. The participant was asked to fill in the same two questionnaires about music and mood, and a new one about Greta, lasting about 2 minutes. The following questionnaire measured the user's personality, and it lasted about 5 minutes. The next minute concerned the collection of some information about the participant, i.e. age, gender, nationality, education, while the last minute was dedicated to the debriefing by the experimenter who came back into the room.

4.3.1 Audio stimuli

The music file provided to the participants consisted of 2 minutes length audio piece in which the music content evoked a funny narration.

Studies by D. Huron (Huron, 2004) identified a set of Humour-Evoking devices in classical music pieces. 5 audio-only stimuli were designed and synthesized by exploiting some of those devices:

- incongruous sounds: they are sounds that do not match the expected music context; for example, during a quite piano performance, a mosquito enters the scene and engages in a music fight with the player;

- metric disruptions: a beat is eliminated or added to a measure; for example, while the 9th Symphony in D-Flat op.125 of L.W. Beethoven is played by a pianist, an extra beat is added at the end of a measure, creating a rhythmic pattern which is not part of the original composition;

- implausible delays: they occur when a music phrase does not end at the expected time; for example, a crescendo performed by an orchestra Implausible Delays: they occur when a music phrase does not resolve in the expected time; for example, a crescendo performed by an orchestra evoking the conclusion of a music peace repeats for an unexpected number of times or with a exaggerated long delay;

- Excessive Repetition: is consists of a repetition of a music passage for a number of times out of the music norm; for example, the ring-tone of a mobile phones is repeated continuously becoming faster and faster;

- Incongruous Quotation: a well-known music piece is mixed with another music piece of an non-congruent style; for example, a quick cavalry charge phrase is mixed with a slow and quiet classical music piece.

The 5 pieces were ranked according to their perceived funniness by asking four listeners to provide a continuous rating of the funniness they perceived in the music, on a slider from 0 (not funny at all) to 1 (very funny). The funniness value was sampled at 20 Hz and stored in a file. The mean funniness curve across listeners was computed and normalized by the maximum funniness value among the participants. Finally, the area under the mean funniness curve of each stimulus was computed and normalized by the stimulus length to obtain a single descriptive value of the funniness. The 2 most funny stimuli have been selected for the experiment, and their difference between their funniness levels was negligible.

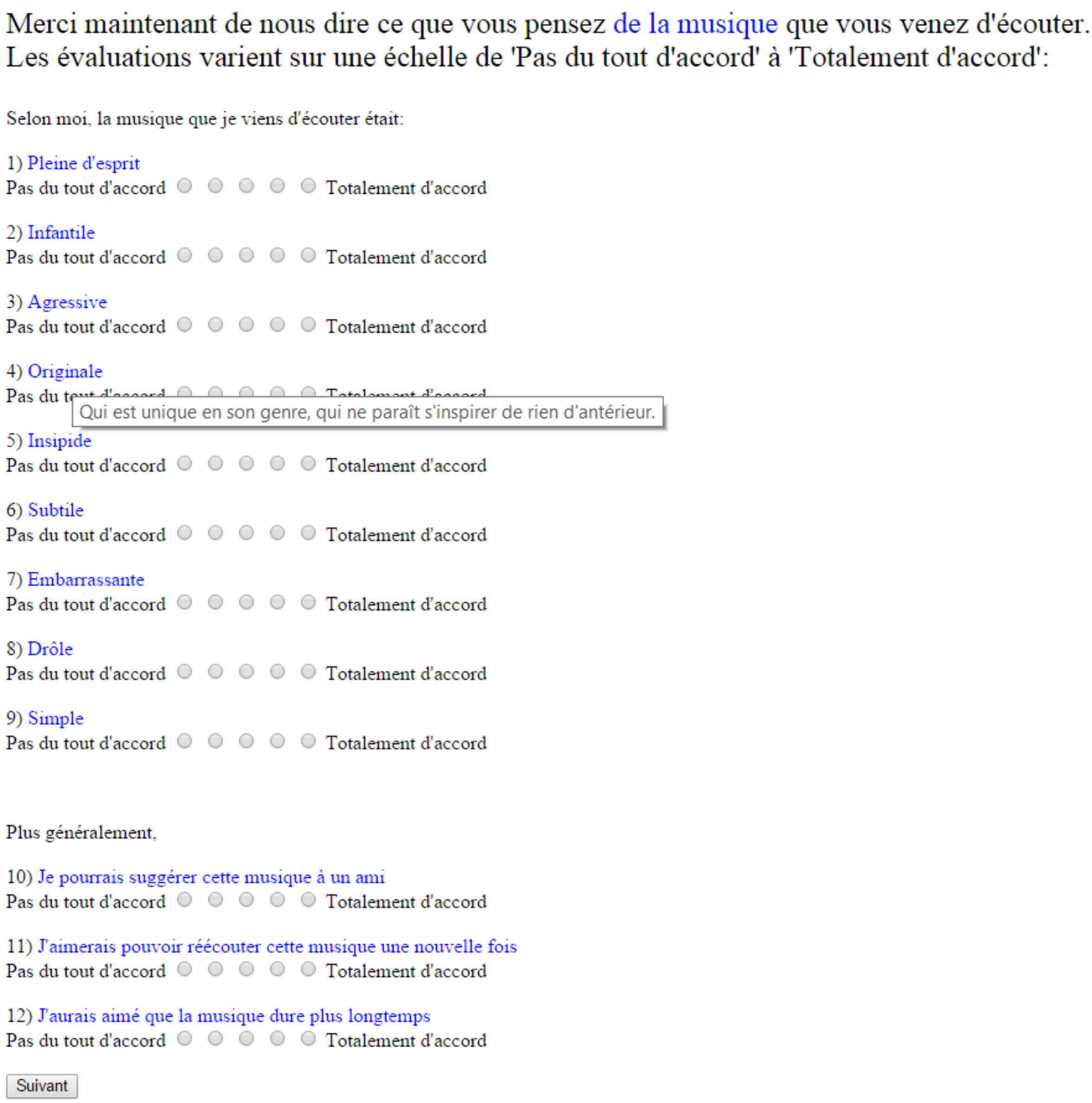
- Stimulus 1 (A1): a pianist is quietly playing the 9. Sinfonie in d-Moll op. 125 of L.W. Beethoven; a mosquito enters the scene and interrupts him multiple times; the pianist becomes upset and tries to hit the mosquito but he hits and break the piano; the mosquito survives to the attack and plays the Sinfonie while flying;

- Stimulus 2 (A2): an orchestra is playing a music piece of Mozart in a concert hall when an Indian tribe enters the room while singing and dancing; the audience reacts by asking them to shut up; the interruptions becomes more and more frequent and loud and the audience start to appreciate them more than the music performance.

4.3.2 Questionnaires

After each listening, participants were asked to fill up some questionnaires: after the first listening, only q_audio and q_funniness, while after the listening with the presence of Greta, q_greta was added after the first two questionnaires. Before demographics information, q_perso was provided to the participants. For a description of the questionnaires see 4.4.2.

The questionnaires were administered via a HTML interface, each questionnaire being showed in a different web page. An example of one of these web pages is showed in Figure 4.2 At the beginning of each questionnaire, a brief instruction about the task was always showed, highlighting the topic which the questions referred to (music, mood, Greta). In case of the presence of adjectives, they were accompanied by a brief definition accessible by the user by scrolling the mouse cursor on the term. In order to avoid any other misunderstanding about the meaning of the questions, the questionnaires were translated in French; regarding q_perso, where the meaning of the adjectives could be a bit subtle, the original English version of the terms was kept inside parenthesis, so that people who spoke English quite well could get the core meaning of the adjectives that could have been partially lost during the translation. The participants were asked to rate their level of agreement on a 5-points Likert scale, except for the questionnaire about personality traits, that included 9-points ranging from "*I do not agree at all*" to "*I totally agree*". When the user completed the questionnaire, he/she could go on by clicking on the button "Continue"; no blank answers were allowed, thus if the user clicked the button without answering all the questions, a warning appeared on the screen and it was no possible to continue.



Merci maintenant de nous dire ce que vous pensez de la musique que vous venez d'écouter. Les évaluations varient sur une échelle de 'Pas du tout d'accord' à 'Totalement d'accord':

Selon moi, la musique que je viens d'écouter était:

1) Pleine d'esprit
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

2) Infantile
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

3) Agressive
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

4) Originale
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord
Qui est unique en son genre, qui ne paraît s'inspirer de rien d'antérieur.

5) Insipide
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

6) Subtile
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

7) Embarrassante
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

8) Drôle
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

9) Simple
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

Plus généralement,

10) Je pourrais suggérer cette musique à un ami
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

11) J'aimerais pouvoir réécouter cette musique une nouvelle fois
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

12) J'aurais aimé que la musique dure plus longtemps
Pas du tout d'accord ○ ○ ○ ○ ○ Totalement d'accord

Suivant

Figure 4.2: An example of the web page showing one of the questionnaires (in this case, q_audio). For item 4 you can see the definition available when scrolling the mouse cursor on the adjective.

During the passage between one page to another, user's answers were transmitted to a Mysql database, containing one table for each questionnaire. The columns of the tables were labelled with:

- ID, relative to the specific table, automatic incrementing in each line;
- "id_subj", unique for each subject, so that it is possible to find the lines corresponding to one subjects, across different tables and different conditions;
- "greta", a number corresponding to the level of factor Copying, thus having two possible values: 1 (no_Copying) and 2 (yes_copying);
- "audio", corresponding to the audio file presented to the participant, thus having two possible values: 1 (A1) and 2 (A2);
- "system", (except for q_greta, where this information does not exists since the participants answered to it only once), corresponding to the level of factor Agent, thus having two possible values: 1 (no_Agent), 2 (yes_Agent);
- "q1","q2","q3",... "qn", containing the scores of the questions.

4.4 Experimental design

The factorial design of an experiment is one of the most important phases, since it determines the following analysis of the data. After several discussions about how to distribute participants between the conditions, we came out with the definitive design, the Independent Variables (IVs) and the Dependent Variables (DVs).

4.4.1 Independent Variables

The variables manipulated during the experiment consisted in one within subjects and one between subjects factor; they are schematised in Figure 4.3.

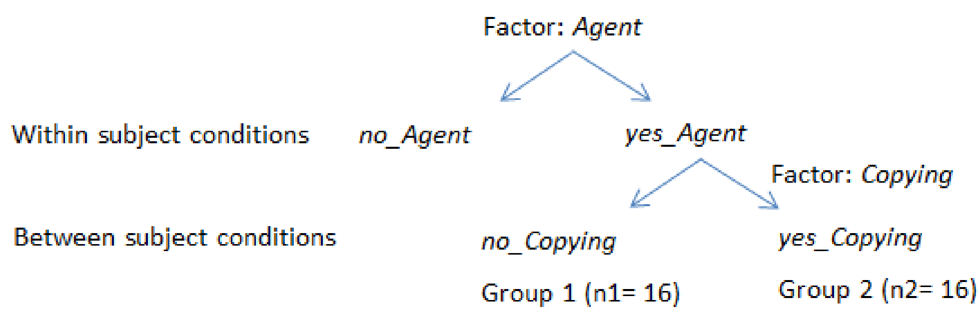


Figure 4.3: The experimental design: *Agent* is the within subject factor and *Copying* is the between subject factor. *Copying* is present only in *yes_Agent* condition: it is related to the virtual agent's behaviour. Each experimental group experienced both *no_Agent* and *yes_Agent* condition.

4.4.1.1 Agent

The within subject factor concerned the presence of the virtual agent during the listening. It had 2 levels:

(i) *no_Agent*, when the participant listened to the music alone and the big screen was black. The choice of not showing a neutral Greta (i.e., in resting condition, without laughing) was made since the main purpose of the questionnaire after this condition was to assess the participant's evaluation of the music, without any external influence;

(ii) *yes_Agent*, when Greta was showed in the big screen, and it listened to the music together with the user.

The order of the two levels of Agent remained *no_Agent*, *yes_Agent* for all the participants: if the participant listened to the music for the first time with Greta, then the following disappearance of the agent could cause disappointment and thus become a confounding variable influencing the following ratings.

4.4.1.2 Copying

The second factor was between subjects and it relied to the level *yes_Agent* of the previous factor. Indeed, when Greta laughed, she could show different behaviours: in *no_Copying* conditions, the agent laughed at prefixed times, with prefixed intensity; in *yes_Copying* conditions, the agent laughed at prefixed times but modulating the intensity according to user's parameters (see 3.3).

The hypothesis of using factor Copying as a within subjects factor, thus assigning each participant to both Copying conditions, was discarded, since by repeating the listening for the third time could introduce some noising effects. For example, the familiarity of the music could positively influence users' ratings, like it has been found to occur for funny videos (crowdsourcing), or, on the other hand, a too long experiment could annoy the participant, and thus negatively influence his/her perception of the experience.

On the whole, the participants listened the music twice, first alone (*no_Agent* condition), then with Greta (*yes_Agent* condition); they were divided into two experimental groups, one of them having Greta laughing independently of user's behaviours (*no_Copying* condition) and the other group having Greta expressing copying behaviour (*yes_Copying* behaviour).

4.4.2 Dependent Variables

The dependent variables were measured through the questionnaires; every DV relied on one of the main hypotheses. All the questionnaires were built upon well-established questionnaires, adding some changes in order to better adapt the questions to our purposes, and used Likert scales.

4.4.2.1 Music Funniness

Participants' perception of music funniness was measured after each listening. The questionnaire aimed to assess this variable was mainly inspired by Ruch and Rath work (Ruch & Rath, 1993). In their study the authors, in order to ask people to evaluate humorous material, used a list of 9 attributes of the stimuli, including: witty, childish, aggressive, original, tasteless, subtle, embarrassing, funny, simple. In our questionnaire, we exploited these adjectives, and we added three additional items adapted from (Prada & Paiva, 2009). The major changes concerned the substitution of the word "games" with "music", and the inversion of the polarity of one sentence from negative to positive; the final sentences used in the questionnaire are: "I would suggest the music to a friend of mine", "I would like to listen to this music again", "I would like that the music lasted more". These last 3 items were aimed at assessing the likeability of the music in an indirect way, while the first nine adjectives directly referred to the music's qualities. In order to avoid any misunderstanding about the meaning of the questions, the questionnaires were translated in French. Hence, the final version of the questionnaire, called q_audio, contained the following items:

- Direct evaluation:
    1. "Pleine d'esprit",
    2. "Infantile",
    3. "Agressive",
    4. "Originale",
    5. "Insipide",
    6. "Subtile",
    7. "Embarrassante",
    8. "Drôle",
    9. "Simple";
- Indirect evaluation:
    1. "Je pourrais suggérer cette musique à un ami",
    2. "J'aimerais pouvoir réécouter cette musique une nouvelle fois",
    3. "J'aurais aimé que la musique dure plus longtemps".

4.4.2.2 Participants' Mood

Users' perception of the experience was investigated by measuring their mood after each listening. The questionnaire, called q_mood, consisted of the 8 attributes of subject's feeling state used by Ruch and Rath (Ruch & Rath, 1993), in the same study described before: exhilarated, bored, activated, indignant, puzzled, angered, amused, and unstimulated. This questionnaire was assigned to the participants at the beginning of the study, in order to assess their initial mood, so that we could discard those people with an extreme mood, since it could influence the following ratings. The French version of q_mood included the following items:

1. "Euphorique",
2. "Ennuyé",

3. "Dynamique",
4. "Indigné",
5. "Perplexe",
6. "En colère",
7. "Amusé",
8. "Amorphe".

4.4.2.3 Spatial presence, social presence and believability

These DVs were measured only after the second listening, since they referred to Greta's characteristics. The questionnaire, called q_greta, was composed by 12 items: some of these items were adapted from Temple Presence Inventory by Lombard (Lombard, Ditton, & Weinstein, 2009), others from Bailenson et al. questionnaire (J. N. Bailenson et al., 2003) and some questions were created by us to better focus on our interests.

The main adaptation we made concerned the use of sentences instead of questions, and the substitution of very general terms with those related to our study, for instance "Greta" instead of "the person", "office" instead of "place".

The final version of the questionnaire, translated in French, contained the following items:

1. "J'ai l'impression d'avoir partagé le même bureau que Greta", adapted from TPI,
2. "Durant l'expérience, je pouvais distinguer les expressions faciales de Greta", from TPI,
3. "Durant l'expérience, je pouvais distinguer le comportement non-verbal de Greta", from TPI,
4. "J'ai senti que je pouvais interagir avec Greta", adapted from TPI,
5. "est crédible", along with the following two sentences, it was created by us, in order to assess the believability,
6. "rit de manière appropriée",
7. "est spontanée",
8. "J'ai eu l'impression que Greta était consciente de ma présence", adapted from Bailenson et al. questionnaire,
9. "J'ai perçu Greta comme un avatar numérique, pas comme une personne réelle", adapted from Bailenson et al. questionnaire,
10. "J'ai eu l'impression que Greta et moi avions partagé le même plaisir durant l'expérience", created by us in order to assess the believability of Greta while laughing,

11. "J'étais principalement concentré sur la musique", along with the following one, it was created by us in order to investigate user's attention, and thus Greta's presence,

12. "J'étais principalement concentré sur Greta".

4.4.2.4 Personality Traits

Finally, the personality of the participants was assessed through a short version of Mini-Markers by Sauciers (Saucier, 1994), further reduced by selecting 3 out of the original 5 traits used by Sauciers, since we tried to let the experiment less boring than possible. Hence, we focused on extroversion, agreeableness and neuroticism, as the traits with more probability to be correlated with sense of humour, the first two in a positive way, while the latter in a negative way. The total number of items was 24, 8 for each trait, 4 positive and 4 negative.

4.5 Hypotheses

We considered three main hypotheses:

- H1: the presence of laughing Greta affects user's perception of the funniness of the music. In particular, we expect the music funniness to increase when the laughing agent is present than when the user listens to the music alone, and to increase more when Greta is in *yes_Copying* conditions than *no_Copying* condition;

- H2: the presence of copying behaviour in Greta, in terms of laughing in a way related to user's laughter intensity, affects user's perception of the experience. In particular, we expect user's mood to be more positive when Greta laughs with an intensity copying those of the user, that when Greta laughs without considering the user;

- H3: the presence of copying behaviour in Greta affects her social presence, spatial presence and believability. In particular, we expect an improvement of these characteristics when Greta's behaviour reflects the user's one.

4.6 Data Analysis

We collected data from 32 participants, 19 males and 13 females, 12 with an age between 18-25 (37.5%), 16 with an age between 26-35 (50%), 3 between 36-45 (9.36%) and 1 between 46-55 (3.13%). The 75% of the participants were French, and the rest correctly understood French. Regarding their education, there were 23 graduated students, 1 undergraduated, 4 of them had a PhD and 4 a High School degree. 20 (62.5%) of them could play an instrument.

4.6.1 Preprocessing

The data were preprocessed by applying some operations on the tables: first of all, the ratings of items referring to negative adjectives have been inverted, by selecting the columns corresponding to those questions and applying a function which transformed the score x into $(maxscore+1)-x$. Then, the columns containing information about the factors were deleted, and a vector was created, where each element represented the sum of the scores of all the questions given by one participant. These vectors were then used in the t-tests.

The preprocessing of the data of q_perso was quite different: we first deleted the columns relative to information about IVs, and we transformed the scores of negative items x into $(maxscore + 1) - x$, here the maximum score was 9 instead of 5 since we used the same Likert scale used by Sauciers. Then, for each personality trait, the vector containing the positions of the adjectives relative to it was used to filter the table, and the output vector contained the average of the scores given by each participant for relative personality trait. Thus, in total we created 3 vectors, one for Extroversion, one for Agreeableness and one for Neuroticism, that were compared to the vectors relative to the ratings of the other questionnaires. Since we run correlation tests, we had to pay attention to the order of the values in the vectors; in this case the column "id_subj" (see 4.3.2 was exploited to select in the vectors of personal traits those positions relative to the scores of the participants assigned to the experimental group under analysis.

4.7 Results

4.7.1 Music funniness

Check for audio. First of all, we checked if the perceived funniness of the different audio files was really the same as stated in 4.3.1. To do this, we considered the vectors of the ratings given by the participants after the first listening, when they were not influenced by other factors. after verifying the normal distribution of the data using Shapiro test (A1: p= 0.9786; A2: p= 0.8677), we run a t-test between the participants who listened to A1 and those who listened to A2. Here is the output of the test:

```
Welch Two Sample t-test

data:  audio1v and audio2v
t = -0.1908, df = 29.405, p-value = 0.425
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
    -Inf 3.45614
sample estimates:
mean of x mean of y
  33.7500   34.1875
```

Being *p-value* » 0.05, the hypothesis of non-significant differences between the two music pieces was not rejected, thus we could follow the initial hypothesis of negligible difference between the two audio files (see 4.3.1). Hence we did not care about possible effects of this variable during the following analyses, i.e. we put the participants listening to A1 and A2 in the same group.

Factor *Agent*. In order to verify H1 4.5, we considered the effect of the factor *Agent*. Also in this case, the Shapiro test confirmed the normal distribution of the data (p(*no_Agent*)= 0.8248; p(*yes_Agent*)= 0.7991), we thus could run a paired t-test. In order to pay attention to any possible effects of factor *Copying*, we first compared participants' ratings after the first and the second listening, and then we did the same, but considered group *yes_Copying* and *no_Copying* separated.

- All the subjects together: we compared the vectors of ratings during condition *no_Agent* and *yes_Agent*:

```
Paired t-test


data:  sistem1v and sistem2v
t = -0.0602, df = 31, p-value = 0.4762
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf 0.8483333
sample estimates:
mean of the differences
               -0.03125
```

  No significant differences were found between the two groups, thus it seems that the presence of the laughing agent did not significantly affect the perception of the music funniness.

- Dividing groups according to Factor *Copying*: we first compared the vectors of ratings during condition *no_Agent* and *yes_Agent*, given by the participants assigned to the condition *no_Copying*. The Shapiro test confirmed the normal distribution of the data (p(*no_Agent*)= 0.8988; p(*yes_Agent*)= 0.9581). The output of t-test was:

```
Paired t-test

data:  sistem1greta1v and sistem2greta1v
t = -0.6086, df = 15, p-value = 0.276
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf 0.9402778
sample estimates:
mean of the differences
                   -0.5
```

  We did the same thing with the vectors of ratings during condition *yes_Copying*; the Shapiro confirmed allowed also in this case for a t-test (p(*no_Agent*)= 0.2093; p(*yes_Agent*)= 0.71):

```
Paired t-test

data:  sistem1greta2v and sistem2greta2v
t = 0.6849, df = 15, p-value = 0.2519
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.6823588        Inf
sample estimates:
mean of the differences
                 0.4375
```

To summarize, no significant differences between the within subjects conditions were found neither when splitting participants according to their between subjects condition.

One possible solution to this unexpected result accounts for possible ambiguities in the questionnaires, or some questions could have been wrongly interpreted by the participants. It is possible that some questions are better than others for the purpose of analysing the DV; we then decided to make two other types of analyses:

1. Dividing the questions in groups/categories, referring to direct items (from 1 to 9, see 4.3.2) and indirect items (from 10 to 12).
   The output for direct questions was:

   ```
   data:  sistem1v and sistem2v
   t = -0.7278, df = 31, p-value = 0.2361
   alternative hypothesis: true difference in means is less than 0
   95 percent confidence interval:
         -Inf 0.4154875
   sample estimates:
   mean of the differences
                  -0.3125
   ```

   While the output for indirect questions was:

   ```
   Paired t-test

   data:  sistem1v and sistem2v
   t = -0.7278, df = 31, p-value = 0.2361
   alternative hypothesis: true difference in means is less than 0
   95 percent confidence interval:
         -Inf 0.4154875
   sample estimates:
   mean of the differences
                  -0.3125
   ```

2. Analysing question-by-question:

   ```
   "Agressive" "0.017" "greater" "greta= 1"
   "Subtile" "0.041" "less" "greta= 1"
   "Je pourrais suggerer cette musique a un ami" "0.041" "greater" "greta= 2"
   ```

   Some adjectives were significant in certain condition of factor *Copying*: the participants who listened to the music with Greta laughing with pre-fixed intensities rated the music less aggressive than it was during the first listening, but more subtle. Unexpectedly, the participants who listened to the music with Greta in copying behaviour were less willing to suggest the music to a friend.

Factor *Copying*. We investigated whether the presence of copying behaviour in the laughing agent had an effect on the perception of music funniness. In this case only data form condition *yes_Agent* were considered, since it is the only level of factor *Agent* where Greta is present during the listening. The Shapiro test confirmed the normal distribution of the data ($p(no\_Copying)= 0.9581$; $p(yes\_Copying)= 0.71$), thus we run a t-test between the vectors of ratings of the participants of the two groups:

```
Welch Two Sample t-test

data:  greta1v and greta2v
t = -0.3011, df = 27.612, p-value = 0.3828
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 3.489952
sample estimates:
mean of x mean of y
   33.625    34.375
```

No significant effects were found also for the factor *Copying*, i.e. no significant differences between the ratings of the participants who listened to the music with Greta laughing with pre-fixed intensities and those who listened to the music with Greta modulating its laughter intensity according to the user's one.

Again, we tried other types of analyses as we did for factor *Agent*.

1. Dividing the questions in groups/categories, referring to direct items (from 1 to 9, see 4.3.2) and indirect items (from 10 to 12).
   The output for direct questions was:

   ```
     Welch Two Sample t-test

   data:  g1s2 and g2s2
   t = -0.5314, df = 28.148, p-value = 0.2997
   alternative hypothesis: true difference in means is less than 0
   95 percent confidence interval:
        -Inf 2.200835
   sample estimates:
   mean of x mean of y
        30.5      31.5
   ```

   When we run the Shapiro test on the distribution of vectors about indirect questions, they resulted not normal distributed ($p(no\_Copying)$= 0.002325; $p(yes\_Copying)$= 0.005407). We thus needed to use a non-parametric test, and we opted for Wilcoxon test:

   ```
   wilcox.test(g1s2,g2s2, alternative="greater")

   Wilcoxon rank sum test with continuity correction

   data:  g1s2 and g2s2
   W = 131.5, p-value = 0.4529
   alternative hypothesis: true location shift is greater than 0
   ```

   Also in this case, no significant results were found.

2. Analysing question-by-question: the only significant result was:

   ```
   "Infantile" "0.026"     "greater"
   ```

   Hence, the participants of the group *yes_Copying* found the music less childish than the participants of the other group did.

4.7.2 Users' mood

Check for audio. First of all, we checked whether the users' mood was not affected by using different audio files. To do this, we considered the vectors of the ratings of q_mood given by the participants after the first listening, when they were not influenced by other factors. After verifying the normal distribution of the data using Shapiro test (A1: p= 0.3868; A2: p= 0.3364), we run a t-test between the participants who listened to A1 and those who listened to A2. Here is the output of the test:

```
Welch Two Sample t-test

data:  audio1v and audio2v
t = 0.5154, df = 29.216, p-value = 0.3051
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -2.152137       Inf
sample estimates:
mean of x mean of y
  27.4375   26.5000
```

Being *p-value* >> 0.05, the hypothesis of non-significant differences between the two music pieces in affecting users' mood was not rejected, thus we could follow the initial hypothesis of negligible difference between the two audio files (see 4.3.1). Hence we did not care about possible effects of this variable during the following analyses, i.e. we put the participants listening to A1 and A2 in the same group.

Factor *Agent*. In order to verify H2 4.5, we considered the effect of the factor *Agent*. Also in this case, the Shapiro test confirmed the normal distribution of the data (p(*no_Agent*)= 0.4323; p(*yes_Agent*)= 0.3062), we thus could run a t-test. In order to pay attention to any possible effects of factor *Copying*, we first compared participants' ratings after the first and the second listening, and then we did the same, but considered group *yes_Copying* and *no_Copying* separated.

- All the subjects together: we compared the vectors of ratings during condition *no_Agent* and *yes_Agent*:

  ```
  Paired t-test

  data:  sistem1v and sistem2v
  t = -0.5522, df = 31, p-value = 0.2924
  alternative hypothesis: true difference in means is less than 0
  95 percent confidence interval:
        -Inf 0.7764427
  sample estimates:
  mean of the differences
                   -0.375
  ```

  No significant differences were found between the two groups, thus it seems that the presence of the laughing agent did not significantly affect user's mood and thus the perception of the experience.

- Dividing groups according to Factor *Copying*: we first compared the vectors of ratings during condition *no_Agent* and *yes_Agent*, given by the participants assigned to the condition *no_Copying*. The Shapiro test confirmed the normal distribution of the data (p(*no_Agent*)$= 0.855$; p(*yes_Agent*)$= 0.3342$). The output of t-test was:

```
Paired t-test

data:  sistem1greta1v and sistem2greta1v
t = -0.5561, df = 15, p-value = 0.2932
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 1.076121
sample estimates:
mean of the differences
                   -0.5
```

We did the same thing with the vectors of ratings during condition *yes_Copying*; the Shapiro test allowed also in this case for a t-test (p(*no_Agent*)$= 0.3827$; p(*yes_Agent*)$= 0.3563$):

```
t.test(s1g2,s2g2,alternative="less",paired=TRUE)

Paired t-test

data:  s1g2 and s2g2
t = -0.2388, df = 15, p-value = 0.4072
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 1.585129
sample estimates:
mean of the differences
                  -0.25
```

To summarize, no significant differences between the within subjects conditions were found neither when splitting participants according to their between subjects condition.

As we did in the previous case for q_audio (see 4.7.1), we decided to make other types of analyses. In this case the division of the questions in groups was not useful because they all came from the same original questionnaire (see 4.3.2), thus they all had the same purpose; we thus carried on the analysis of the single questions:

```
"Dynamique" "0.014"     "less"      "greta= 1"
"Amorphe" "0.028" "less" "greta= 1"
```

These results showed that the users felt more dynamic but at the same time more amorphe after listening to the music with Greta in *no_Copying condition.*

Factor *Copying*. We investigated whether the presence of copying behaviour in the laughing agent had an effect on the users' mood. In this case only data form condition *yes_Agent* were considered, since it is the only level of factor *Agent* where Greta is present during the listening. The Shapiro test confirmed the normal distribution of the data (p(*no_Copying*)$= 0.3342$; p(*yes_Copying*)$= 0.3563$), thus we run a t-test between the vectors of ratings of the participants of the two groups:

```
t.test(g1s2,g2s2,alternative="less")

Welch Two Sample t-test
```

```
data:  g1s2 and g2s2
t = -0.8136, df = 28.656, p-value = 0.2113
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 1.838026
sample estimates:
mean of x mean of y
  28.5000   30.1875
```

No significant effects were found also for the factor *Copying*, i.e. no significant differences between the ratings of the participants who listened to the music with Greta laughing with pre-fixed intensities and those who listened to the music with Greta modulating its laughter intensity according to the user's one.

Again, we tried other types of analyses, i.e. considering single questions: we did not find any significant result.

4.7.3 Spatial presence, social presence and believability

In order to verify H3 (see 4.5), we compared the ratings of the participants for q_greta only in the condition *yes_Agent*, being the only case where Greta was present during the listening.

Check for audio. We checked also in this case whether the two music pieces had an influence on the DVs, using a t-test (Shapiro test: p(A1)=0.1515; p(A2)=0.7742). The output was:

```
Welch Two Sample t-test

data:  audio1v and audio2v
t = 0.9369, df = 29.993, p-value = 0.1782
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.420342       Inf
sample estimates:
mean of x mean of y
  36.9375   35.1875
```

Being *p-value* >> 0.05, the hypothesis of non-significant differences between the two music pieces in affecting Greta's characteristics was not rejected, thus we could follow the initial hypothesis of negligible difference between the two audio files (see 4.3.1). Hence we did not care about possible effects of this variable during the following analyses, i.e. we put the participants listening to A1 and A2 in the same group.

Factor *Copying*. In order to investigate whether the characteristics of Greta were improved by its copying behvious, we compared the vectors of ratings of the participants for q_greta; the Shapiro test confirmed the normal distribution of the data (p(*no_Copying*)= 0.401; p(*yes_Copying*)= 0.9849), thus we run a t-test:

```
Welch Two Sample t-test

data:  greta1 and greta2
t = -0.2061, df = 28.647, p-value = 0.4191
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 3.624283
sample estimates:
```

```
mean of x mean of y
  35.0625   35.5625
```

Since the results did not show any significant differences between the two groups, we continued with the division of the questions in subgroups referring to spatial presence, social presence and believability.

- Spatial presence: we selected the questions relative to this DV ("J'ai l'impression d'avoir partagé le même bureau que Greta","J'étais principalement concentré sur la musique", "J'étais principalement concentré sur Greta"), and after running the Shapiro test (p(no_Copying)= 0.05955; p(yes_Copying)= 0.03146) we opted for the non-parametric Wilcoxon test:

  ```
  wilcox.test(g1,g2,alternative="less")

  Wilcoxon rank sum test with continuity correction

  data:  g1 and g2
  W = 127, p-value = 0.4923
  alternative hypothesis: true location shift is less than 0
  ```

- Spcial presence: we selected the questions relative to this DV ("J'ai senti que je pouvais interagir avec Greta", "J'ai eu l'impression que Greta était consciente de ma présence", "J'ai eu l'impression que Greta et moi avions partagé le même plaisir durant l'expérience"), and after running the Shapiro test (p(no_Copying)= 0.1633; p(yes_Copying)= 0.5846) we opted for the t-test:

  ```
  Welch Two Sample t-test

  data:  greta1 and greta2
  t = -2.0025, df = 29.906, p-value = 0.02719
  alternative hypothesis: true difference in means is less than 0
  95 percent confidence interval:
         -Inf -0.1713745
  sample estimates:
  mean of x mean of y
      3.875     5.000
  ```

- Believability: we selected the questions relative to this DV ("Durant l'expérience, je pouvais distinguer les expressions faciales de Greta", "Durant l'expérience, je pouvais distinguer le comportement non-verbal de Greta", "est crédible", "rit de manière appropriée", "est spontanée", J'ai perçu Greta comme un avatar numérique, pas comme une personne réelle"), and after running the Shapiro test (p(no_Copying)= 0.9882; p(yes_Copying)= 0.1153) we opted for the t-test:

  ```
  Welch Two Sample t-test

  data:  g1 and g2
  t = 0.2833, df = 29.502, p-value = 0.3894
  alternative hypothesis: true difference in means is greater than 0
  95 percent confidence interval:
   -1.56043      Inf
  sample estimates:
  mean of x mean of y
    21.0625   20.7500
  ```

The results were significant for Spatial presence: the ratings of the participants were higher in the group where Greta laughed copying users' laughter intensity.

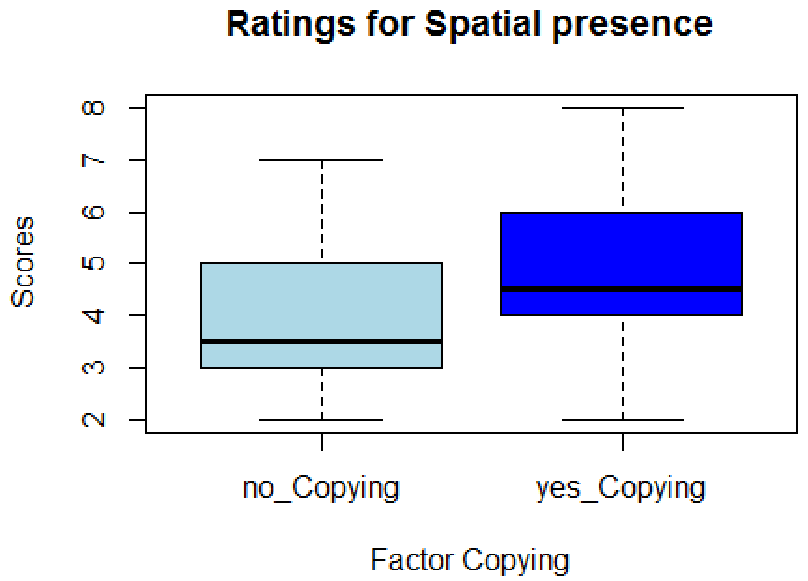


Figure 4.4: The significant difference of ratings about social presence of Greta between *no_Copying* group and *yes_Copying* group.

Then, we analysed the single questions:

```
"Durant l'exp\'erience, je pouvais distinguer le comportement non-verbal de Greta" "0.024" "greater"
"J'ai senti que je pouvais interagir avec Greta" "0.021" "less"
"J'ai eu l'impression que Greta \'etait consciente de ma pr\'esence" "0.006" "less"
```

The questions that resulted as significant were relative to social presence, confirming the previous results. In addition, the non-verbal behaviour of Greta was perceived m when the agent was copying the user's laughter intensity. Thus, it seems that when the agent is more tuned to user's behaviour, its behaviour is less easy to distinguish, maybe because it becomes more natural, while in *no_Copying* condition, where Greta was laughing randomly, its behaviour could have been more strange, and thus more easily noticeable.

4.7.4 Personal Traits

The scores given by participants to q_perso were analysed in order to search correlation with the ratings in the other questionnaires.

Correlations with q_audio. We run correlation tests between each trait and the ratings of participants about music funniness, according to different conditions: the ratings after the first listening; those after the listening with Greta; the difference between ratings after the second and the first listening; the ratings of the *no_Copying* group (after the second listening); the ratings of the *yes_Copying* group (after the second listening). Table 4.1 shows the $p - value$ and rho coefficient of each correlation.

Personal trait \ Ratings		**first list**	**second list**	**diff**	***no_Cop***	***yes_Cop***
Extrov	*p-value*	0.9339	0.8588	0.4960	0.6214	0.3899
	ρ	0.0153	0.0327	-0.1248	-0.1338	0.2307
Agr	*p-value*	0.3953	0.1266	0.1070	0.0605	0.7572
	ρ	0.1555	0.2758	0.2903	0.4791	0.0839
Neurot	*p-value*	0.5994	0.7574	0.0314	0.8321	0.8579
	ρ	-0.0965	0.0568	0.3810	0.0576	-0.0486

Table 4.1: Correlations between personal traits and ratings of questionnaire q_audio.

We found only one significant correlation, between Neuroticism and the difference of ratings after the second and first listening: participants with higher scores for neuroticism evaluated the music as funnier after the listening with Greta. In other words, the effect of Greta on music funniness is quite higher for people with higher levels of neuroticism.

Correlations with q_mood. We run the same correlation tests considering the ratings of participants about their mood. Table 4.2 shows the $p - value$ and rho coefficient of each correlation. No significant correlations were found for these data.

Personal trait \ Ratings		**first list**	**second list**	**diff**	***no_Cop***	***yes_Cop***
Extrov	*p-value*	0.9554	0.5186	0.5697	0.4642	0.8466
	ρ	-0.0103	0.1184	0.1044	0.1972	0.05260
Agr	*p-value*	0.7900	0.7528	0.2944	0.2435	0.6609
	ρ	0.04900	0.0579	0.1912	0.3094	-0.1189
Neurot	*p-value*	0.8169	0.8860	0.5493	0.3445	0.4408
	ρ	0.0426	-0.0264	0.1099	0.2530	-0.2074

Table 4.2: Correlations between personal traits and ratings of questionnaire q_mood.

Correlations with q_greta. Finally, we run correlation tests for the ratings about spatial presence, social presence and believability of Greta. In this case, the ratings of participants concerned: the ratings of all the participants; the ratings of the *no_Copying* group; the ratings of the *yes_Copying* group. Table 4.3 shows the $p - value$ and rho coefficient of each correlation. Again, no significant results were found.

4.8 Discussion

The results did not allow us to confirm our hypotheses. We did not find any significant effect neither of factor *Agent*, nor of factor *Copying* on the ratings of participants about music funniness and their mood.

Ratings / Personal trait		all	*no_Copying*	*yes_Copying*
Extroversion	*p-value*	0.7845	0.5083	0.9935
	ρ	-0.0503	-0.1785	0.0022
Agreeableness	*p-value*	0.174	0.9011	0.131
	ρ	0.2464	-0.0338	0.3941
Neuroticism	*p-value*	0.9048	0.1033	0.1976
	ρ	0.0220	0.4222	-0.3400

Table 4.3: Correlations between personal traits and ratings of questionnaire q_greta.

The only interesting results concern the evaluation of social presence: the groups who listened to the music with Greta when copying user's laughter intensity rated its social presence with higher scores than did the participants who listened to the music with Greta who was laughing randomly. This result confirm the importance for a virtual agent to exploit data from the user and copy user's behaviours, in a direct or indirect way, since this enhance its social presence.

4.8.1 Possible Suggestions

There are many points that, if investigated and improved, could let us hope to find more interesting results in the next months.

First of all, at the end of the experiment we asked participants for some feedback about the experience and the music they listened. One important issue concerned the fact the the music was not so funny as expected from the theories followed to create it. Most of the users smiled rather than laughed, and some of them found the music too childish, or, in some cases, annoying. The lack of significant influence of Greta on user's experience could be due to the weak funniness of our stimuli, that is, the virtual companion could enhance the user's experience, but starting from a stimulus that has a certain level of funniness itself; otherwise, the use of audio stimuli in general could be not so effective in triggering laughter, than videos or verbal puns are. Moreover, many participants revealed that they appreciated the music, but not at a level that made them laugh: they explicitly told us that the presence of a video representing the funny situation occurring in the audio files would ease them to laugh, like if in this way the story was more concrete. They found quite strange to laugh only at an audio stimulus. This observation seems to confirm the relative weakness of audio stimuli with respect to other types of stimuli in triggering laughter. To face this problem, we are planning to analyse the videos captured during the listening, since they might help us to understand which part of the music triggered laughter and thus, was considered as funny. This would allow us to know where to improve the funniness of the different pieces of music, thus we could investigate whether the audio stimulus in general is not effective, or our stimulus was not well designed.

Another future improvement should concern the extension of the sample size, indeed we are planning to continue the study in Rovereto, in order to collect more participants and also make a comparative study between different cultures, French and Italian, to investigate if there is some influence of this factor in the sense of humour and disposition to laugh. If we look at the means of the DVs compared in the various t-tests (see 4.7), we can notice that in most of the cases the mean of the ratings of *yes_Copying* group are higher than those of *no_Copying* group, as well as the mean of the ratings of *yes_Agent* group are higher than those of *no_Agent* group. These are the results that we looked for in order to validate our hypotheses (see 4.5), but they are not statistically significant. If this tendency of the data persisted when testing more participants, they could have the chance to reach significant values.

Another element that we could improve is the copying behaviour of Greta: in addition to dynamically adapting the agent's body movements, we ought to adapt the agent's laughter sounds. Moreover, the agent gaze needs to be enhanced. Our synthesis model does not compute the eyes direction of the agent as our motion capture data on which the data-driven synthesis model is built does not contain such data. By combining data-driven and procedural animation model, the agent will be able to gaze at times at the participants, giving the impression to perceive their the presence. By improving the believability, we might help the participants to feel more involved and thus improve their experience.

Bibliography

Askenasy, J. J. M. (1987). The journal of general psychology. *Brain*, *114*(4), 317–334.

Bailenson, J. N. et al. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, *29*(7), 819–833.

Bailenson, J. & Yee, N. (2005). Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*(10), 814–819.

Berg, D. V. & Van Brockern, S. (1995). Building resilience through humor. *Reclaiming Children and Youth: Journal of Emotional and Behavioral Problems*, *4*(3), 26–29.

Bevacqua, E. et al. (2007). An expressive eca showing complex emotions. In *Proceedings of the aisb annual convention, newcastle, uk* (pp. 208–216).

Bevacqua, E. et al. (2010). Greta: towards an interactive conversational virtual companion. *Artificial Companions in Society: perspectives on the Present and Future*, 143–156.

Borod, J. et al. (1997). Neuropsychological aspects of facial asymmetry during emotional expression: A review of the normal adult literature. *Neuropsychology Review*, *7*(1), 41–60.

Broekema, C. (2011). Motion intensity matching in interaction with a virtual agent. In *14th twente student conference on it.*

Brownell, H. H. & Gardner, H. (1988). Neuropsychological insights into humour.

Buchowski, M., Majchrzak, K., Blomquist, K., Chen, K., Byrne, D., & Bachorowski, J. (2007). Energy expenditure of genuine laughter. *International journal of obesity*, *31*(1), 131–137.

Buschmeier, H., Bergmann, K., & Kopp, S. (2010). Adaptive expressiveness: virtual conversational agents that can align to their interaction partner. In *Proceedings of the 9th international conference on autonomous agents and multiagent systems: volume 1-volume 1* (pp. 91–98). International Foundation for Autonomous Agents and Multiagent Systems.

Carlson, P. & Peterson, R. L. (1995). What is humor and why is it important? *Journal of Emotional and Behavioral Problems*, *4*(3), 6–12.

Cassell, J. et al. (2001). More than just a pretty face: conversational protocols and the affordances of embodiment. *Knowledge-Based Systems*, *14*(1), 55–64.

Castellano, G. et al. (2012). Expressive copying behavior for social agents: a perceptual analysis. *IEEE Transactions on Systems Man and Cybernetics-Part A-Systems and Humans*, *42*(3), 776.

Chartrand, T. & Bargh, J. (1994). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Chatbots.org. (2015a). Chatbot. Retrieved from https://www.chatbots.org/chatbot/

Chatbots.org. (2015b). Eliza. Retrieved from https://www.chatbots.org/chatbot/eliza

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In *Darwin and facial expression: a century of research in review* (pp. 11–89).

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Marry.

Dearborn, G. V. (1900). The nature of the smile and laugh. *Science*, 851–856.

Ding, Y. et al. (2014). Laughter animation synthesis. In *Aamas'14* (pp. 773–780). Paris, France.

Douglas-Cowie, E. et al. (2003). Emotional speech: towards a new generation of databases. *Speech communication*, *40*(1), 33–60.

Douglas-Cowie, E. et al. (2007). The humaine database: addressing the collection and annotation of naturalistic and induced emotional data. In *Affective computing and intelligent interaction* (pp. 488–500). Springer.

Drack, P., Huber, T., & Ruch, W. (2009). *The apex of happy laughter. a facs-study with actors.*

Ekman, P. [P], Friesen, W., & Hager, J. (2002). The new facial action coding system (facs). *Research Nexus division of Network Information Research Corporation.*

Foot, H. C. & Chapman, A. J. (1976). The social responsiveness of young children in humorous situations. *Humour and laughter: Theory, research, and applications*, 187–214.

Fry, W. F. (2002). Humor and the brain: a selective review. *Humor*, *3*(15), 305–333.

Goel, V. & Dolan, R. J. (2001). The functional anatomy of humor: segregating cognitive and affective components. *Nature Neuroscience*, *4*(3), 237–238.

Gruner, C. (1978). *Understanding laughter: the workings of wit and humor*. Nelson-Hall.

Hauck, W. E. & Thomas, J. W. (1972). The relationship of humor to intelligence, creativity, and intentional and incidental learning. *The Journal of Experimental Educational*, 52–55.

Hayashi, T., Urayama, O., Hori, M., Sakamoto, S., Nasir, U. M., Iwanaga, S., . . . Murakami, K. (2007). Laughter modulates prorenin receptor gene expression in patients with type 2 diabetes. *Journal of psychosomatic research*, *62*(6), 703–706.

Heylen, D. et al. (2001). Embodied agents in virtual environments: the aveiro project.

Huron, D. (2004). Music-engendered laughter: an analysis of humor devices in pdq bach. In *Proceedings of the 8th international conference on music perception and cognition. evanton, illinois* (pp. 700–704).

Johnson, W. L. & Rickel, J. (1997). Steve: an animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin*, *8*(1-4), 16–21.

Kimata, H. (2004). Effect of viewing a humorous vs. nonhumorous film on bronchial responsiveness in patients with bronchial asthma. *Physiology & behavior*, *81*(4), 681–684.

Krenn, B. et al. (2011). Embodied conversational characters: representation formats for multimodal communicative behaviours. In *Emotion-oriented systems* (pp. 389–415).

Lombard, M., Ditton, T., & Weinstein, L. (2009). Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence* (pp. 1–15).

MacGhee, P. E. & Pistolesi, E. (1979). *Humor: its origin and development.* WH Freeman and Company.

Mancini, M. et al. (2008). Greta: a saiba compliant eca system. In *Troisiéme workshop sur les agents conversationnels animés.*

Mancini, M. et al. (2012). Computing and evaluating the body laughter index. In *Human behavior understanding* (pp. 90–98).

Mancini, M. et al. (2014). Laugh when you are winning. In *Innovative and creative developments in multimodal interaction systems* (pp. 50–79).

Marsi, E. & van Rooden, F. (2007). Expressing uncertainty with a talking head in a multimodal question-answering system. In *Mog 2007 workshop on multimodal output generation* (p. 105).

Martin, R. A. (2007). The psychobiology of humor and laughter. In *The psychology of humor: an integrative approach.* (Pp. 153–189). Academic Press.

McDougall, W. (1922). Why do we laugh? *Scribners*, *71*, 359–363.

McKeown, G. et al. (2012). Ilhaire laughter database. In *Proceedings of the lrec workshop on corpora for research on emotion sentiment and social signals (es 2012). european language resources association (elra), istanbul.*

Mehrabian, A. (2007). *Nonverbal Communication.* Aldine.

Meyer, M. et al. (2007). How the brain laughs: comparative evidence from behavioral, electrophysiological and neuroimaging studies in human and monkey. *Behavioural brain research*, 345–260.

Morkes, J. et al. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: a direct test of srct theory. *Human-Computer Interaction*, *14*(4), 395–435.

Morreal, J. (1983). A new theory. In *Taking laughter seriously.* SUNY Press.

Nagaoka, C., Komori, M., & Yoshikawa, S. (2007). Embodied synchrony in conversation. In *Conversational informatics* (pp. 331–351).

Neurology. (2012). Periaqueductal gray.

Nielsen, J. (1994). *Usability engineering.*

Niewiadomski, R. [R.] & Pelachaud, C. (2012). Towards multimodal expression of laughter. In *Intelligent virtual agents* (pp. 231–244).

Niewiadomski, R. [Radoslaw] et al. (2009). Greta: an interactive expressive eca system. In *Proceedings of the 8th international conference on autonomous agents and multiagent systems-volume 2* (pp. 1399–1400). International Foundation for Autonomous Agents and Multiagent Systems.

Nijholt, A. (2002). Embodied agents: a new impetus to humor research.

Pauchet, A. & Sabouret, N. (2012). *Embodied conversational agents and affective computing.* 14th European Agent Systems Summer School. Retrieved from http://www.slideshare.net/easss2012/t7-embodied-conversational-agents-and-affective-computing

Picard, R. (1997). Affective computing. 1997.

Platt, T. et al. (2012). Experimental standards in research on ai and humor when considering psychology. In *Aaai fall symposium: artificial intelligence of humor*.

Platt, T. & Forabosco, G. (2013). Gelotophobia: the fear of being laughed at. *Humor and health promotion*, 229–252.

Poeck, K. (1969). Pathophysiology of emotional disorders associated with brain damage.

Poggi, I. et al. (2005). Greta. a believable embodied conversational agent. In *Multimodal intelligent information presentation* (pp. 3–25). Springer.

Pollio, H. R. et al. (1972). Humor, laughter, and smiling: some preliminary observations of funny behaviors. *The psychology of humor*, 211–239.

Prada, R. & Paiva, A. (2009). Teaming up humans with autonomous synthetic characters. *Artificial Intelligence*, *173*(1), 80–103.

Proyer, R. T. et al. (2008). The fear of being laughed at (gelotophobia) and personality. *Anuarul Institutului de Istorie" George Baritiu" din Cluj-Napoca-Seria HUMANISTICA*, (6), 53–68.

Radcliffe-Brown, A. (1952). *Structure and functions in primitive society: essays and addresses.* Free Press.

Reeves, B. & Nass, C. (1996). *How people treat computers, television, and new media like real people and places.* CSLI Publications and Cambridge university press.

Roether, C., Omlor, L., & Giese, M. (2008). Lateral asymmetry of bodily emotion expression. *Current Biology*, *18*(8), 329–330.

Ruch, W. et al. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor: International Journal of Humor Research*, *27*(1), 23–45.

Ruch, W. & Ekman, P. [Paul]. (2001). The expressive pattern of laughter. *Emotion, qualia, and consciousness*, 426–433.

Ruch, W. & Proyer, R. T. (2008). The fear of being laughed at: individual and group differences in gelotophobia.

Ruch, W. & Rath, S. (1993). The nature of humor appreciation: toward an integration of perception of stimulus properties and affective experience. *Humor: International Journal of Humor Research.*

Saucier, G. (1994). Mini-markers: a brief version of goldberg's unipolar big-five markers. *Journal of personality assessment*, *63*(3), 506–516.

Shneiderman, B. (1986). *Designing the user interface-strategies for effective human-computer interaction.*

Slater, D. (2000). *Interactive animated pedagogical agents* (PhD thesis, Stanford University).

Szabo, A. (2003). The acute effects of humor and exercise on mood and anxiety. *Journal of Leisure Research*, *35*(2), 152.

Trouvain, J. (2003, August). Segmenting phonetic units in laughter. In U. A. de Barcelona (Ed.), *Proceedings of the 15th international congress of phonetic sciences.* Barcelona.

Urbain, J. et al. (2009). Avlaughtercycle: an audiovisual laughing machine. In *Proceedings of the 5th international summer workshop on multimodal interfaces* (pp. 79–87).

Urbain, J. et al. (2010). The avlaughtercycle database. In *Lrec.*

Urbain, J. et al. (2013a). Laugh machine. *Proceedings eNTERFACE*, *12*, 13–34.

Urbain, J., Niewiadomski, R., Mancini, M., Griffin, H., Çakmak, H., Ach, L., & Volpe, G. (2013b). Multimodal analysis of laughter for an interactive system. In *Intelligent technologies for interactive entertainment* (pp. 183–192).

Vilhjálmsson, H. et al. (2007). The behavior markup language: recent developments and challenges. In *Intelligent virtual agents* (pp. 99–111).

Wikipedia. (2014a). Embodied agent — Wikipedia, the free encyclopedia. [Online; accessed 29-April-2014]. Retrieved from http://en.wikipedia.org/wiki/Embodied_agent

Wikipedia. (2014b). Facial electromyography — Wikipedia, the free encyclopedia. [Online; accessed 22-December-2014]. Retrieved from http://en.wikipedia.org/wiki/Facial%5C_electromyography

Wikipedia. (2015a). Avatar (computing)— Wikipedia, the free encyclopedia. [Online; accessed 8-March-2015]. Retrieved from http://en.wikipedia.org/wiki/Avatar_(computing)

Wikipedia. (2015b). Chatterbot — Wikipedia, the free encyclopedia. [Online; accessed 3-February-2015]. Retrieved from http://en.wikipedia.org/wiki/Chatterbot

Wikipedia. (2015c). Eliza — Wikipedia, the free encyclopedia. [Online; accessed 2-March-2015]. Retrieved from http://en.wikipedia.org/wiki/ELIZA

Wikipedia. (2015d). Intelligent agent — Wikipedia, the free encyclopedia. [Online; accessed 2-March-2015]. Retrieved from http://en.wikipedia.org/wiki/Intelligent_agent

Wikipedia. (2015e). Non-player character — Wikipedia, the free encyclopedia. [Online; accessed 11-March-2015]. Retrieved from http://en.wikipedia.org/wiki/Non-player%5C_character

Wild, B. et al. (2003). Neural correlates of laughter and humour. *Brain*, *126*(10), 2121–2138.

Wilkins, J. & Eisenbraun, A. J. (2002). Humor theories and the physiological benefits of laughter. *Holistic nursing practice*, *23*(6), 349–354.

Yamao, Y. et al. (2014). Neural correlates of mirth and laughter: a direct electrical stimulation study. *Cortex*.