how social perception affects sentence
choices in the dialog tree??

Interpersonal stance recognition using non-verbal signal on several time windows

Mathieu Chollet, Magalie Ochs, Catherine Pelachaud⋆
{mchollet, mochs, cpelachaud}@telecom-paristech.fr

⋆CNRS-LTCI, Telecom Paristech
75013 PARIS – FRANCE

Abstract:
We present a computational model for interpreting non-verbal signals of an user during an interaction with a computer in order to obtain a representation of his interpersonal stance. Our model analyses signals and reactions to signals in their immediate context, as well as features of signal production patterns and reaction patterns on different time windows : reaction time, dialogue turn, conversation topic. In this paper, we propose a first model parameterized using data obtained from the literature on the expressions of stances through interpersonal behavior.

Keywords: Interpersonal stance, non-verbal behavior interpretation, Social Signal Processing

1 Introduction

The last two decades have seen a surge of interest in the field of Human-Computer Interaction for the introduction of Embodied Conversational Agents (ECA) in various application domains, such as interactive storytelling[8], virtual learning environments[10][16], healthy behaviour promotion [7], or museum guides[14]. One of the major reasons behind this strong movement is that some studies found that using ECAs could potentially improve the experience of human-computer interaction, by making learning activities easier to follow[**?**] or by enhancing the degree of trust users had in relationship with their computer[19].

Moreover, several researchers have recently focused on Social Signal Processing (SSP). The objective of SSP is to allow computers to sense social signals, such as boredom, politeness, or interpersonal stances[20].

Synthesis on the one hand, analysis on the other, it is clear that both domains are very complimentary and could benefit from each other. Interacting with an ECA is yet the form of Human-Computer Interaction that is closest to a real Human-Human interaction. it could thus be hypothetised that humans display the same social signals interacting with ECAs as they do with their human peers. Then, ECAs would prove a very interesting way to gather data for Social Signal Processing research.

Social Signal Processing is also a very valuable tool for research on ECA behavior. One of the ways to make ECAs more believable when they interact with an user, is to give them the capability to adapt themselves to that user's interpersonal stance. For example, in the context or a virtual learning environment, it would be useful for a virtual teacher to detect when a learning user feels embarassed, as it would allow the ECA to adapt its behavior and its teaching strategy.

The work presented in this paper is part of TARDIS, a FP7 funded project whose objective is to help with inclusion of the increasing number of young Europeans not in employment, education or training. The vision of the TARDIS project is to give young people a tool to train their social skills : a serious game of job interviews simulation, that will help them improve their chances of getting a job. One of the research challenges TARDIS aims at tackling is the recognition of interpersonal stances. Indeed, in a job interview, recruiters try to assess social skills of candidates by judging their interpersonal dispositions and social attitudes : to improve their performance, candidates thus have to adapt their behavior by strategically adopting the best stance at every moment. For example, when discussing management skills, a job candidate wants to appear dominant, and when discussing team-working abilities, a job candidate wants to appear friendly and not too dominant.

In the TARDIS platform, the recruiter will be a virtual recruiter enacted by an ECA and the recognition of social signals will be automated, using sensors such as webcams and microphones. Therefore, the TARDIS project is a perfect example of the combination of Social Signal Processing and Embodied Conversation Agents : we have to detect interpersonal stances in real-time, we have to know how the virtual recruiter should react through its own interpersonal stances, and both of these matters should be adressed keeping the context of the interaction in mind.

However, there is still significant progress to be made in Social Signal Processing to account for long-term social stance recognition. As Pantic recently states in [15],

> *« despite a significant progress in automatic recognition of audiovisual behavioural cues underlying the manifestation of various social signals, most of the present approaches to machine analysis of human behaviour are neither multimodal, nor context-sensitive, nor suitable for handling longer time scales. In turn, most of the social signal recognition methods reported so far are single-modal, contextinsensitive and unable to handle long-time recordings of the target phenomena. »*

This paper proposes a model for interpersonal stance recognition aiming at tackling these issues, namely analysing multimodal social signals on several temporal scales, taking the context into account. For the temporal issue, we propose to use different time windows of analysis. Relevant signals are not necessarily the same in every time window. To take into account the context, as a first step, we analyse the signal reactions of the user to the ECA's signals and verbal phrases. We consider signals from different modalities, however multimodality (i.e. combinations of signals mean more than just a juxtaposition of signals) will only be considered in ulterior versions of the model.

The rest of the paper is organised as follows. Section 2 presents existing works in perception of interpersonal stance, affect recognition and ECA interaction related to our goal. In section 3 we give out definitions for the notions of interpersonal stance, social and verbal signals, reactions, features, and time windows. Section 4 then proposes a first version of our model, considering a few modalities only, and based on psychology litterature. Section 5 discusses future work, with an emphasis on how we plan to tune our model using annotated corpora of videos of conversations.

2 Related Work

2.1 Perception of interpersonal stances in agents

In order to study perception of social attitudes, some authors have generated variations of ECAs behavior displays and asked users to rate how dominant they perceived the agent.

Fukayama *et al.*[11] have proposed a gaze movement model for embodied agents based on three parameters : the amount of gaze directed at the interlocutor, the mean duration of gaze directed at the user, and the gaze points while averting gaze. They found that variation of these three parameters allowed their agents to convey different impressions of dominance and friendliness to users.

Bee *et al.*[4] studied the relationsip between signals of several different modalities and the perception of social dominance of a virtual character. They analysed the relationship between different facial expressions of emotions (joy, fear, anger, surprise, disgust, neutral), different head and gaze orientations, and how users perceived the dominance of the resulting face. They showed that variations of gaze and head orientations don't always have the same effects depending on the displayed emotion. In [5], they looked at the relationship between the head, gaze orientations and parameters of sentence generation (more or less extraverted or agreeable), and found that neither the verbal or non-verbal modality truly dominates in rating a character's personality.

In [2], Arya *et al.* study the effect of facial expressions of virtual characters on the perception of their personality, by displaying videos of a virtual character displaying a specific expression with a certain speed. They then asked them to choose an adjective from a list that suited best the expression. The list of adjectives only contained words characteristic to a specific region of the interpersonal circumplex, a bidimensional representation of interpersonal stances (See §3.1 for more details). As a result, they were able to link the facial expressions to specific points on the interpersonal circumplex, thus providing a direct mapping from behavior to personality.

2.2 Multimodal social signals recognition

Wagner *et al.* proposed a framework, SSI, for Social Signal Interpretation in the context of online recognition systems. It supports input from a variety of sensors and is equipped with algorithms to perform multimodal fusion. In a sample application [21], SSI was plugged in with the « Alfred » ECA and the agent would try to mirror the user's emotional state by using the approriate facial expressions. The recognition of the emotion state consistes of merging audio and video signals to get a dimensional representation of the user's affect, in terms of pleasure and arousal. However, the estimation of the user's affective states is done *a posteriori*, and perceptions of interpersonal stance can be different if we are directly interacting with other people or if we are assessing interpersonal stances as outside spectators.

Few attempts have been made for estimation of the most dominant person in a small group meeting [17] [13]. However those works are offline methods for groups of people and might not be applicable in our setting, i.e. real-time human-machine interaction. They still bear some insight as to what nonverbal signals are the most relevant in assessing perception of dominance. The strongest cues were found in most cases to simply be the total speaking time of the participants.

2.3 Interactions using social signals

Some recent systems have used social signals as inputs to drive an interaction. In [9], Cavazza *et al.* use emotional speech to drive an interactive narrative taking place within an adaptation of Flaubert's *Madame Bovary*. The emotional features of the user's voice that are recognised by the system are used as part of the scenario planning : the interpretation of the signal is dependent of the context. One of the main advantages of their approach is that the interaction is driven without any verbal recognition or semantic interpretation, which allows for completely free speech from the users, while still allowing for variability in the scenarii.

As part of the SEMAINE project, an integrated platform of Sensitive Artificial Listeners was developped. It consists of affect recognition modules (video, audio inputs) that are fed to a listener model that was developped by Bevacqua *et al.* [6]. The different listeners have different personality and signal production rules : the enthusiastic and cheerful Poppy will very often mimic the user's behavior and produce a lot of backchannels, while the antagonist Spike will display social signals that are contrary to those that the user displays.

Existing works in analysis of social signals in interaction have for now been for the biggest part focusing on recognition and use of emotions. In contrast, interpersonal stances have not received much attention. However, endowing agents with the capability of detecting interpersonal stances, and colouring their behavior with interpersonal stances, would prove very useful in the TARDIS project and applications of similar nature. This paper provides a model for interpersonal stance recognition in such contexts.

3 Definitions

3.1 Interpersonal stance

In [18], Scherer provides a specification for the attributes that differentiate the types of affective phenomena : emotions, moods, attitudes, preferences, affect dispositions, and interpersonal stances. For him, the specificity of interpersonal stances is that

> *« it is characteristic of an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons, coloring the interpersonal exchange in that situation (e.g. being polite, distant, cold, warm, supportive, contemptuous). »*

Attitudes towards others are mapped by Arygle [1] on two dimensions : Dominance/Submissive and Friendly/Hostile, in line with works on interpersonal behavior that consistently found these two axes to account for most of the nonverbal variations, such as the Interpersonal Circumplex proposed by Wiggins [22] (See Fig.1).

To represent the user's interpersonal stance, we choose to use two dimensions, based on Argyle's attitude dimensions and Wiggin's interpersonal circumplex axes : friendliness (also called warmth or affiliation) and dominance (also called agency). The interpersonal stance of an agent U at a time t (virtual or human) is formally represented as follows :

$$Stance_U(t) = \{Dom_U(t), Frnd_U(t)\}$$

with $Dom_U, Frnd_U \in [-1, 1]$ representing respectively the dominance and the friendliness expressed by an user U. The more Dom_U (resp. $Frnd_U$) is close to 1, the more dominant (resp. friendly) the user. The more Dom_U (resp. $Frnd_U$) is close to -1, the more submissive (resp. hostile) the user.

3.2 Social signals

The notion of signal varies a lot depending on the domain of study. In Social Signal Processing (SSP), a general consensus over the definition of

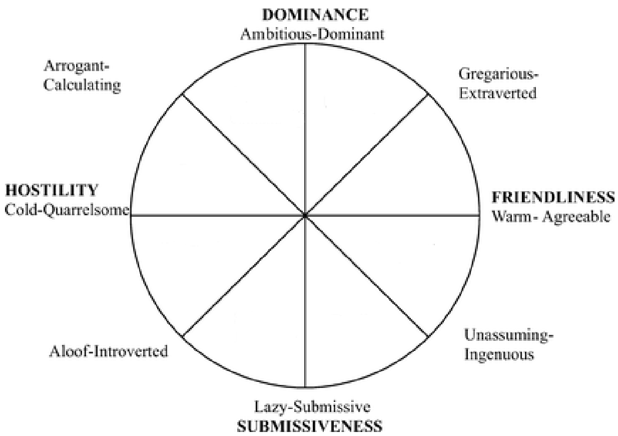

FIGURE 1 – An interpersonal Circumplex with prototypical interpersonal stances every 45°

what is a social signal is hard to find. In [20], Vinciarelli *et al.* review the relevant notions and propose the following definiton :

> *« A Social signal is a communicative or informative signal that, either directly or indirectly, provides information about social facts, namely social interactions, social emotions, social attitudes, or social relations. »*

In our case, we rely on external software that sends messages when it detects non-verbal signals. We thus consider that a non-verbal signal is an input message characterized by a starting time t_{start}, an end time t_{end}, and a non-verbal body modality (e.g. gaze, facial expression, voice). For each of these modalities, a number of additional relevant variables are defined : for instance, a signal from the gaze modality also contains two angles, one used to know the direction of gaze aversion (angle around the head front axis), and one to know how much gaze is averted (angle between the head front axis and the gaze direction axis).

In an immediate post-processing step, we classify these non-verbal messages in specific types depending on the values of their variables. For instance, any gaze signal averted from more than 5° is classified as a *gazeAway* signal. As a first step, we consider the following modalities and types of signals, because they were related to interpersonal stances in previous works :

- Gaze :
  - *gazeFront (GzFr)*, when the angle between the gaze direction and the front axis of the head is less than 5°
  - *gazeAway (GzAw)*, when the angle is higher than 5°
- Head orientation :
  - *headFront (HeFr)*,when the angle between the head front axis and the person-interlocutor axis is less than 5°
  - *headUp (HeUp)*, when this angle is more than 5° and the user is looking upwards
  - *headDown (HeDo)*, when this angle is more than 5° and the user is looking downwards
  - *headSide (HeSi)*, when this angle is more than 5° and the user is looking sideways
- Facial expressions : *smile (Smle)*, when the lip corners are turned up [1]
- Voice : *isSpeaking (Spk)*, when the audio signal exceed a certain threshold [1]

3.3 Verbal signals

In the context of the TARDIS platform, there is no speech recognition. However, in the context of an interaction with a conversational agent, we know in advance the sentences uttered by the ECA, and the types of these sentences (praise, criticisms, etc...) should be considered to analyse the user's reactions. Indeed, a signal may be interpeted totally differently depending on what it is expressed in reaction to. For instance, smiling can be considered as arrogance when criticized, whereas smiling when praised expressed pride.

To categorize the ECA's sentences, we use the typology proposed by Bales in his Interaction Process Analysis (IPA) Theory[3]. We reduce it to three categories : sentences can either be socio-emotional positive (SEpos), socio-emotional negative (SEneg), or task-oriented (TO).

3.4 Reactions

The signals we described in previous section can be either in isolation, that is they are displayed spontaneously by the person with no direct relation do the other interactant's behavior. But signals can be expressed in direct reaction to the other interactant's behavior.

A reaction R consists of the signal expressed $s_{reaction}$ in reaction to the signal s_{origin}. When

1. In a later version, we will define more types using other variables such as voice pitch, smile intensity. For the sake of simplicity we consider very few types.

Type	Categories
Socio-Emotional Positive *SEpos*	1. Seems friendly 2. Dramatizes 3. Agrees.
Attempted Answers *A*	4. Gives suggestion 5. Gives opinion 6. Gives information
Questions *Q*	7. Asks for information 8. Asks for opinion 9. Asks for suggestion
Socio-Emotional Negative *SEneg*	10. Disagrees 11. Shows tension 12. Seems unfriendly

TABLE 1 – Bales IPA categories[3]

an isolated signal is noted simply as : $s_{isolated}$, a reaction R is noted in the following manner.

$$R = s_{reaction} \leftarrow s_{origin}$$

Determining if a signal is isolated or in reaction to another signal is a hard problem. In a simplifying assumption, we consider that if a signal s_A is sent by the person A, then any signal s_B sent by B during the $\Delta_{REACTION}$ time window of length δ (See next section) is a reaction to s_A. In a more formal notation, we have :

$$\begin{aligned} if \quad & s_{reaction} : t_{start} \in [s_{origin} : t_{start} + \delta, \\ & \qquad s_{origin} : t_{end} + \delta] \\ then \quad & s_{reaction} \leftarrow s_{origin} \end{aligned}$$

In the next section, we explore more particularly the features of signal and reaction production that are relevant in assessing interpersonal stance.

3.5 Features

The analysis of affective phenomena has to be done at different temporal levels depending on the type of phenomena considered. An emotion is a strong local phenomenum (even though it can last), and considering signals on a short time window is enough to detect them. Interpersonal stance, on the other hand, is inferred on longer temporal scales, by analysing reccuring tendencies in behavior, and not only single signals occurences at a particular point in time.

For every type of signals (e.g. smiles or gaze aversion in reaction to criticism) we define relevant features to assess users' interpersonal stance. These features are used to evaluate more long-term characteristics of the signal productions : for instance, if the user responds to a smile of the agent by another smile, it can be considered as a sign of friendliness but it is not sufficient to infer that the user has a friendly stance. On the other hand, the amount of smile reactions the user has produced in reaction to agent smiles on a longer time scale may give us more reliable information.

In this version, we consider the following features for each non-verbal signal types.

- Amount of this signal type (e.g. percentage of gazeFront, or number of smiles)
- Mean duration of this signal type (in seconds)
- Amount of reactions of this type after *Socio-emotional positive* answers
- Amount of reactions of this type after *Socio-emotional negative* answers
- Amount of reactions of this type after a *Smile* signal

3.6 Time windows

Non-verbal signals give out cues about the mental state of the person that displays them. For instance, seeing a person suddenly frown their brows, clench their fists and raise their voice energy are cues that hint this person is angry at this precise moment.

However, as Scherer points out [18], all kinds of affect don't happen in the same span of time. For instance, emotions have a very short duration, and to assess a person's emotion, one should only look at this person's very recent displays of emotion in their non verbal behavior. For moods, one has to look at another person's non-verbal behavior on a longer time span. And it might get even longer to get a good sense of someone's personality.

Therefore, to recognize interpersonal stances, we should consider doing analysis of non-verbal behavior on different time spans. For this purpose, we define three time windows of analysis.

Reaction window. The $\Delta_{REACTION}$ window is very short, starting from a signal S start time and lasting for a small constant δ. δ is the length of the time frame in which we can consider that an interlocutor's signal is still in reaction to S. In that time window, there is no specific feature extraction, however every user signal detected in this window is labeled as a reaction to S.

Dialogue turn window. The Δ_{TURN} window is the current turn of dialogue. Its starting time is the point where the current speaker has started talking, and it lasts until the other interactant takes the floor. In this time window, we are interested in signals occurences, reaction occurences, and some patterns.

The following features are extracted in the Δ_{TURN} time window :

– Amount of $gazeFront$, $headFront$, $headUp$, $headDown$, $headSide$
– Occurence (or absence) of a smile in this turn ($true$ or $false$)
– If the agent uttered a sentence of type $SEpos$ (resp. $SEneg$)
  – Absence of reaction
  – Aversion of gaze (occurence of $gazeAway$)
  – Aversion of head (occurence of $headUp$, $headSide$, or $headDown$)
  – Occurence of smile

Dialogue topic window. The Δ_{TOPIC} window starts when a new topic is being discussed. To represent the topic discussed during the dialog, we use a dialogue model based on hierarchical task networks. We consider that this window starts when a new top-level task (e.g. greetings, discuss resume, discuss job experience) starts, and ends when another top-level task starts. This is a much longer time window and we are more interested in features of signal and reactions patterns here.

The following features are extracted in the Δ_{TOPIC} time window :

– Amount of $gazeFront$, $headFront$, $headUp$, $headDown$, $headSide$
– Mean duration of $gazeFront$, $gazeAway$, $headFront$, $headUp$, $headDown$, $headSide$, $isSpeaking$, $smile$
– Amount of agent $SEpos$ (resp. $SEneg$) sentences with no reaction, $smile$ reaction, $gazeAway$ reaction, $headUp$, $headDown$, or $headSide$ reaction

4 Computation of interpersonal stance

4.1 Problem definition

In essence, our problem is to find the relationship between two variables, dominance (Dom_U) and friendliness($Frnd_U$), and a set of input variables $X(t) = x_i(t)$, where each x_i is one of the n features of signals and reactions patterns defined in the previous section. We want to find the functions D and F such that $Dom_U(t) = D(X(t))$ and $Frnd_U(t) = F(X(t))$.

In a simplifying assumption, we suppose that the input variables are independent, which allows us to split the problem of finding the functions D and F into smaller problems of finding the relationship between an input variable and dominance and friendliness independently of the others. Specifically, we suppose that $D(X(t)) = \sum_{i=1}^{n} Dw_i * D_i(x_i(t))$, where each D_i models the relationship between the variable x_i and dominance independently of other signals, and Dw_i is a weighting factor. The same supposition is made for friendliness, so we have $F(X(t)) = \sum_{i=1}^{n} Fw_i * F_i(x_i(t))$

4.2 Relationships between input variables and stance

In our case, the relationship between dominance or friendliness and the non-verbal behavior is not always close to linear. For instance, studies on gaze and mutual gaze [11] have shown that a medium to high amount of gaze are rated neutral or slightly positively on the friendliness scale, but a low or very high amount of gaze are rated as negative on the same scale.

The psychological litterature provides good insights about the general properties of the relationship between interpersonal stance and nonverbal behavior. However, precise mappings between these signal patterns and the interpersonal stance dimensions are hard to find.

Considering this, it is hard to make strong assumptions concerning the precise shape of the relationship between patterns of non-verbal signals and perception of interpersonal stance (e.g. logarithmic vs exponential...). Then, in order to use this knowledge while refraining from making too strong assumptions on the shape of these functions, we decide to adopt piecewise linear function shapes. That is, we consider that the functions that map features to dominance and friendliness are linear on intervals (See Fig. 2). For instance, friendliness is increasing with amount of gaze on the interval [0%,50%], it is highest and nearly constant in the interval [50%,80%] and decreasing on the interval [80%,100%].

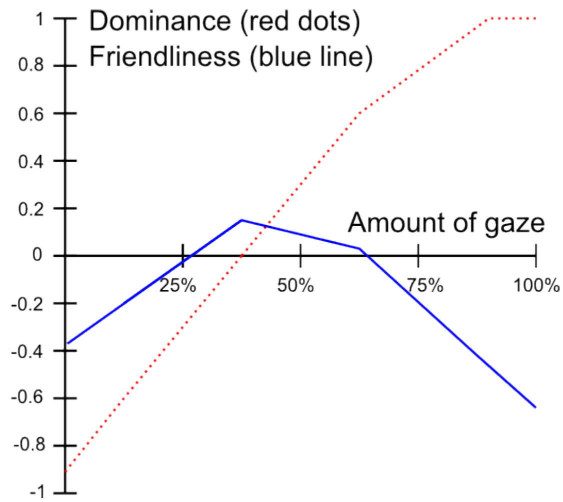

FIGURE 2 – Example of piece-wise linear function of dominance (red) and friendliness (blue), based on [11].

4.3 Tuning the weights of stance equations

Once the shape of the functions D_i and F_i have been found, the only thing remaining is to adjust the corresponding weights (the Dw_i and Fw_i) to reflect the contribution of every input variable with respect to the overall interpersonal stance perception.

However, as we haven't gathered data yet, we rely once again on psychological knowledge to tune the first version of our model. More specifically, in [12], Gifford computes correlations between specific non-verbal modalities occurences and perceptions of interpersonal stance. The more correlated is the non-verbal behavior with dominance or friendliness, the strongest weight we assign to it.

Once those two steps are done, the model can be used online to compute, at every time t, the value of the user's interpersonal stance.

5 Conclusion

We have presented a computational model for interpersonal stance recognition. This model takes into account the interactional nature of a conversation, by considering that a spontaneous signal gives different information than a signal in reaction to an other person's behaviour. It also analyses behavior on different temporal patterns, by using several time windows.

In this first version we tuned the system using data from psychological litterature. In a next step, we plan on learning the parameters of the model using real data. For this, we will use videos annotated with occurences of non-verbal signals and interpersonal stances ratings.

We also want to tackle the issue of multimodality : the combination of non-verbal signals can mean something different than just the sum of them. For instance, clenching one's fist can mean anger, and smiling can indicate friendliness. However, the combination of both is used when celebrating success.

6 Acknowledgement

The research leading to this paper has received funding from the European Union Information Society and Media Seventh Framework Programme FP7-ICT-2011-7 under grant agreement 288578.

References

[1] M. Argyle. *Bodily Communication*. University paperbacks. Methuen, 1988.

[2] Ali Arya, Lisa N. Jefferies, James T. Enns, and Steve DiPaola. Facial actions as visual cues for personality. *Computer Animation and Virtual Worlds*, 17(3-4) :371–382, 2006.

[3] R.F. Bales. *A Set of Categories for the Analysis of Small Group Interaction.Channels of Communication in Small Groups*. Bobbs-Merrill, 1950.

[4] Nikolaus Bee, Stefan Franke, and Elisabeth André. Relations between facial display, eye gaze and head tilt : Dominance perception variations of virtual agents.

[5] Nikolaus Bee, Colin Pollock, Elisabeth André, and Marilyn Walker. Bossy or Wimpy : Expressing Social Dominance by Combining Gaze and Linguistic Behaviors. In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271, Berlin, Heidelberg, 2010. Springer Berlin / Heidelberg.

[6] Elisabetta Bevacqua, Etienne De Sevin, Julia Hyniewska, Sylwia, and Catherine Pelachaud. A listener model : introducing personality traits. *Journal on Multimodal User Interfaces, special issue Interacting ECAs*, page 12, 2012.

[7] Timothy W. Bickmore and Rosalind W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2) :293–327, June 2005.

[8] Marc Cavazza, Fred Charles, and Steven J. Mead. Character-based interactive storytelling. *IEEE Intelligent Systems*, 17(4) :17–24, July 2002.

[9] Marc Cavazza, David Pizzi, Fred Charles, Thurid Vogt, and Elisabeth André. Emotional input for character-based interactive storytelling. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '09, pages 313–320, Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems.

[10] Mark Core, David Traum, H. Chad Lane, William Swartout, Stacy Marsella, Jonathan Gratch, and Michael Van Lent. Teaching negotiation skills through practice and reflection with virtual humans. simulation. *SIMULATION*, 82 :685–701, 2006.

[11] Atsushi Fukayama, Takehiko Ohno, Naoki Mukawa, Minako Sawaki, and Norihiro Hagita. Messages embedded in gaze of interface agents — impression management with agent's gaze. *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, (4) :41, 2002.

[12] Robert Gifford. Mapping nonverbal behavior on the interpersonal circle. *Journal of Personality and Social Psychology*, v61(n2) :p279(10), 1991-08-01. table Means, correlations, and a number of reversals for nonverbal behaviors.

[13] Dinesh Babu Jayagopi, Hayley Hung, Chuohao Yeo, and Daniel Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *Trans. Audio, Speech and Lang. Proc.*, 17(3) :501–513, March 2009.

[14] Stefan Kopp, Lars Gesellensetter, Nicole C. Krämer, and Ipke Wachsmuth. Lecture notes in computer science. chapter A conversational agent as museum guide : design and evaluation of a real-world application, pages 329–343. Springer-Verlag, London, UK, UK, 2005.

[15] M. Pantic, R. Cowie, F. D'Errico, D. Heylen, M. Mehu, C. Pelachaud, I. Poggi, M. Schröder, and A. Vinciarelli. *Social Signal Processing : The Research Agenda*, pages 511–538. Springer, London, 2011.

[16] Jeff Rickel and W. Lewis Johnson. Animated agents for procedural training in virtual reality : Perception, cognition, and motor control. *APPLIED ARTIFICIAL INTELLIGENCE*, 13 :343–382, 1998.

[17] R. J. Rienks and D. Heylen. Automatic dominance detection in meetings using easily detectable features. In *Proc. Workshop Mach. Learn. Multimodal Interaction (MLMI)*, Edinburgh, U.K, Jul. 2005.

[18] K. R. Scherer. What are emotions ? and how can they be measured ? *Social Science Information*, 44 :695–729, 2005.

[19] Susanne van Mulken, Elisabeth André, and Jochen Müller. The persona effect : How substantial is it ?, 1998.

[20] Alessandro Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder. Bridging the gap between social animal and unsocial machine : A survey of social signal processing. *IEEE Transactions on Affective Computing*, 3 :69–87, 2012.

[21] Johannes Wagner, Florian Lingenfelser, Nikolaus Bee, and Elisabeth André. Social signal interpretation (ssi). *KI - Künstliche Intelligenz*, 25 :251–256, 2011. 10.1007/s13218-011-0115-x.

[22] J.S. Wiggins. *Paradigms of Personality Assessment*. Guilford Publications, 2003.