Interpersonal stance recognition using non-verbal signals 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 nonverbal signals of a user during an interaction with a virtual character in order to obtain a representation of his interpersonal stance. Our model starts, on the one hand, from the analysis of multimodal signals. On the other hand, it takes into account the temporal patterns of the interactants behaviors. That is it 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 : signal reaction, sentence reaction, conversation topic, whole interaction. 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 [17], healthy behaviour promotion [7], or museum guides[13]. One of the major reasons behind this strong movement is that some studies found that using ECAs improved the experience of human-computer interaction, by making learning activities easier to follow [17] or by enhancing the degree of trust users had in relationship with their computer [21].

Moreover, several researchers have recently focused on Social Signal Processing (SSP). The objective of SSP is to allow computers to recognize social information, such as boredom, politeness, or interpersonal stances [22].

One of the ways to make ECAs more believable when they interact with a 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. Social Signal Processing provides valuable tools to build ECAs that are capable of reacting to a user.

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 we have, in the context of TARDIS, 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 appropriate stance at every moment. For example, when discussing management skills, a job candidate may want to appear dominant, and when discussing team-working abilities, a job candidate may want 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 the user's interpersonal stance in real-time and we have to know how the virtual recruiter should react to it. One way to react is to express a particular interpersonal stance : for instance the virtual recruiter may decide to express coldness to a dominant user. The context of the interaction should be considered both for detecting the user's interpersonal stance and for deciding which interpersonal stance the agent should express.

In Social Signal Processing, there is still significant progress to be made for 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. Signals do not necessarily convey the same information in every time window. For instance, a smiling person might be interpreted as friendly, whereas someone who smiles in response to criticism can be seen as arrogant. We consider signals from different modalities [1].

The rest of the paper is organised as follows. Section 2 presents related works on perception of ECA's interpersonal stances, multimodal social signal recognition and human-ECA interaction driven by users' affect recognition. In section 3, we introduce the definitions we use in our model for the notions of interpersonal stance, social and verbal signals, reactions, features, and time windows. In Section 4 is proposed a first version of our model based on Social and Human Sciences studies. In Section 5, we conclude and discuss future works.

2   Related Work

2.1   Perception of interpersonal stances in agents

In order to study perception of social attitudes, some researchers generated different ECAs behavior expressions showing different interpersonal stances. Users then had to rate how they perceived the agent.

For instance, Fukayama *et al.* [10] 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 expressed on several different modalities and the perception of social dominance of an ECA. 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 do not 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 both the verbal and non-verbal modalities have an effect on the perception of the ECA's dominance.

In [2], Arya *et al.* studied the effect of facial expressions of ECAs on the perception of their interpersonal stance, by displaying videos of an ECA displaying a specific expression with a certain speed. They then asked users 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 interpersonal stance.

These works highlight the fact that ECAs are capable to convey different interpersonal stances through non-verbal behavior. In the next section, we present existing works on multimodal social signal recognition.

2.2   Multimodal social signals recognition

Wagner *et al.* [23] proposed a framework called *SSI* (Social Signal Interpretation), for designing online recognition systems. This framework supports inputs from a variety of sensors and is equipped with algorithms to perform multimodal fusion. In a sample application [23], *SSI* was plugged in with the *Alfred* agent [4]. The agent mirrors the user's emotional state by using

---

1. However multimodality (i.e. combinations of signals meaning more than just a juxtaposition of signals) will only be considered in ulterior versions of the model.

appropriate facial expressions. The recognition of the user's emotional state is based on both audio and video signals, and yields a dimensional representation of the user's affect in terms of pleasure and arousal [23].

Few attempts have been made for estimation of the most dominant person in a small group meeting [18] [12]. 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.

As shown in the above presented works, systems for affect recognition have been proposed. However, the recognition of interpersonal stances has yet to be attempted during a real-time interaction. In the next section, we present works where users' social signals are used to drive the interaction with ECAs.

2.3 Interactions using social signals

Some recent systems have used users' social signals to drive human-ECA interactions. Cavazza *et al.* [9] use emotional speech to drive an interactive narrative taking place within an adaptation of Flaubert's *Madame Bovary*. Emotional features of the user's voice are recognised by the system and are used as part of the scenario planning. 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 [20], 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 developped by Bevacqua *et al.* [6]. Different listeners with different personalities were implemented. The rules of signal production are dependent on the ECA personality : the enthusiastic and cheerful Poppy will often mimic the user's behavior and will produce a lot of backchannels, while the hostile Spike will display social signals that are contrary to those expressed by the user.

For now, most works in analysis of social si-

gnals in interaction have focused on recognition of users' 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 enable enhancing human-ECA interactions. This paper provides a computational model for users' interpersonal stance recognition in human-ECA interaction.

In the next section, before diving into the details of our model, we introduce definitions for some of its central notions.

3 Definitions

3.1 Interpersonal stance

In [19], 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 Argyle [1] on two dimensions : Dominant/Submissive and Friendly/Hostile. This is in line with works on interpersonal behavior that consistently found these two axes accounted for most of the non-verbal behavior variations, such as the Interpersonal Circumplex proposed by Wiggins [24] (See Fig.1).

Based on Argyle's attitude dimensions and Wiggin's interpersonal circumplex axes, we propose to use two dimensions to represent users' interpersonal stances : *friendliness* (also called warmth or affiliation) and *dominance* (also called agency).

A user can express a general stance in the whole interaction, a stance when discussing a certain topic, and a stance in reaction to specific signals or sentences. For instance, in a job interview, a candidate might be embarassed by a question on a specific skill he does not

have, even though he might have appeared as confident on the overall topic of discussing his skills. Therefore we want to recognize the interpersonal stance a user U expresses in reaction to a signal ($SignalStance_U$), to a sentence ($SentenceStance_U$), during the time a specific topic is discussed ($TopicStance_U$), and on a whole interaction ($InteractionStance_U$).

For an agent U (virtual or human), all of these stances are formally represented as a combination of a *dominance* and a *friendliness* :

$$Stance_U = \{Dom_U, Frnd_U\}$$

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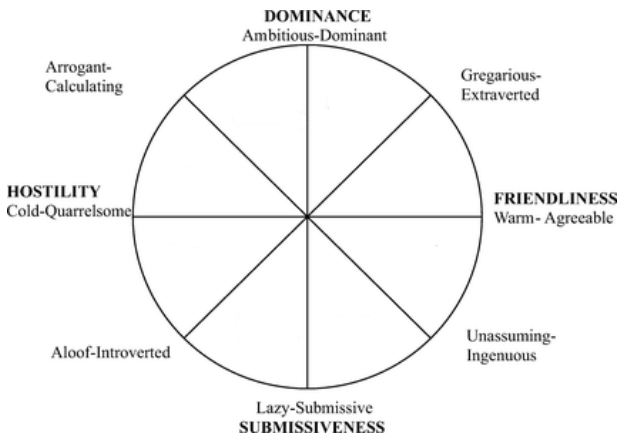

FIGURE 1 – An interpersonal Circumplex with prototypical interpersonal stances every 45°

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 what is a social signal is hard to find. Based on a study of the different definitions of a signal in different disciplines, Vinciarelli *et al.* [22] propose the following definition of social signals :

> « *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 our model, we propose to classify these non-verbal messages in specific types depending on the values of their variables. For instance, when the user is not looking at the ECA, the user's gaze signal is classified as a *gazeAway* signal. Research showed that gaze [10], head orientation [4], smiles [14] and speech [18] were found to reflect users' interpersonal stances. Although other non-verbal signals can be related to it, we choose as a first step to consider those modalities in our model :

– Gaze :
  – *gazeFront*, when the user's gaze is directed at the ECA
  – *gazeAway*, when the user's gaze is not directed at the ECA
– Head orientation :
  – *headFront*, when the user's head is directed at the ECA
  – *headUp*, when the user's head is directed upwards
  – *headDown*, when the user's head is directed downwards
  – *headSide*, when the user's head is directed sideways
– Facial expressions : *smile*, when the user is smiling [2]
– Voice : *speech*, when the user is speaking [2]

3.3 Verbal signals

In the context of the TARDIS platform, keyword spotting is implemented but this is not sufficient for speech recognition and understanding of the user. However, in the context of an interaction with a virtual recruiter, we know in advance the sentences it utters as they are formally represented in a dialogue manager, Disco [16].

---

2. 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.

The types of these sentences (praise, criticisms, etc...) should be considered to analyse the user's reactions. Indeed, a signal may be interpreted differently depending in reaction to what it is expressed. For instance, a smile can be considered as arrogance when it is expressed in reaction to criticism, whereas a smile in response to praise might be interpreted as pride.

To categorize the ECA's sentences, we base ourselves on the typology proposed by Bales in his Interaction Process Analysis (IPA) Theory [3] (See Table 1). In our model, we consider three categories : sentences can either be socio-emotional positive (noted SEpos), socio-emotional negative (noted SEneg), or task-oriented (noted TO) when the sentence is a question or an answer [3].

Type	Categories
Socio-Emotional Positive *SEpos*	1. Seems friendly 2. Tension release 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]

3.4 Reactions

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

An isolated signal is noted simply as : $s_{isolated}$. A reaction R is another kind of signal that is expressed in reaction to another signal s_{origin}. This relationship is noted in the following manner :

$$R \leftarrow s_{origin}$$

Determining if a signal is isolated or in reaction to another signal is a hard problem. In a sim-

plifying assumption, we consider that if a signal s_{origin} is sent by the person A, then any signal $s_{reaction}$ sent by B during the $\Delta_{REACTION}$ time window of length δ (See next section) is a reaction to s_{origin}. In a more formal notation, we have :

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

In the next section, we explore more particularly the features of signals and of reactions produced 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 might be 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) we define relevant features to assess users' interpersonal stance. These features are used to evaluate more stance 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 general 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 gives us additional information.

In our model, we consider the following kinds of features :

– amount of a signal type (e.g. percentage of $gazeFront$, or number of smiles)
– mean duration of a signal type (in seconds)
– amount of reactions of a type after the agent utters a *Socio-emotional positive* sentence
– amount of reactions of a type after the agent utters a *Socio-emotional negative* sentence
– amount of reactions of a type after the agent utters a *Task-Oriented* sentence

3. We propose to simplify Bales' typology by merging the Questions and Attempted Answers categories in a single Task-Oriented category

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 [19], 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 a person's non-verbal behavior on a longer time span. It might get even longer to get a good sense of someone's interpersonal stance.

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

Signal reaction window. The signal reaction window, noted Δ_{SIGNAL}, aims at detecting reactions from the user to *non-verbal signals* of the ECA in order to compute the user's $SignalStance_U$ for this signal. For a signal S expressed by the ECA, the Δ_{SIGNAL} window is very short, starting from the 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 this time window, we are interested in the types of user signals expressed in reaction to S.

Sentence reaction window. The $\Delta_{SENTENCE}$ window aims at detecting reactions from the user to *sentences* uttered by the ECA in order to compute the user's $SentenceStance_U$ for this signal. Its starting time is the point where the ECA starts the sentence, and it lasts until either the user takes the floor and finishes talking or the agent starts another sentence. In this time window, we are interested in the types of user signals expressed in reaction to the ECA's sentence types : socio-emotional positive, socio-emotional negative, or task-oriented (See Section 3.3).

Dialogue topic window. Users can have a specific interpersonal stance regarding a particular discussed topic. For instance, someone can be embarassed when discussing personal matters in an official context. Therefore, we consider a specific window for every topic discussed, and we use it to compute the user's $TopicStance_U$. 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 begins when a new top-level task (e.g. greetings, discuss resume, discuss job experience) starts, and ends when another top-level task starts. The features used in this time window are described in the Section 3.5.

Interaction window. Finally, the user's $InteractionStance_U$ represents the global stance that a user has expressed through an interaction. It is computed on the $\Delta_{INTERACTION}$ time window, that spans from the beginning of the interaction to its end. The features used in this time window are the same as the ones used to compute the user's stance towards the topic, and are described in Section 3.5.

This section has introduced the notions that are used in our model. In the next section, we present how stance is computed.

4 Computation of interpersonal stance

4.1 Problem definition

Our model aims at computing the stance of a user in reaction to a specific signal ($SignalStance$), to a sentence ($SentenceStance$), within a discussion topic ($TopicStance$) or in an entire interaction ($InteractionStance$).

In essence, our problem is to compute for one kind of stance, both the *dominance* (Dom_U) and *friendliness* ($Frnd_U$), from a set of input variables $X = \{x_i; 1 \leq i \leq n\}$, where each x_i is one of the n features used for that stance (see Section 3.6). We want to find the functions D and F such that $Dom_U = D(X)$ and $Frnd_U = F(X)$.

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 $D(X) = \sum_{i=1}^{n} Dw_i * D_i(x_i)$, 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

$$F(X) = \sum_{i=1}^{n} Fw_i * F_i(x_i)$$

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 [10] 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 non-verbal 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 piece-wise linear function shapes. That is, we consider that the functions that map features to dominance and friendliness are linear on intervals. Using data from reports such as [10], we can find appropriate intervals and slopes for these functions. For instance, in [10], dominance is rated from -1 to -1 for amounts of gaze in $\{25\%, 50\%, 75\%, 100\%\}$. We can then draw a piece-wise linear function that passes through those points (See Fig. 2).

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 stance perception.
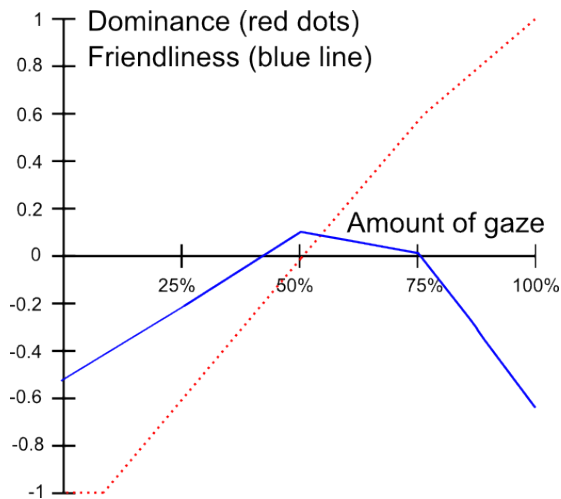

FIGURE 2 – Example of piece-wise linear function of dominance (red dots) and friendliness (blue line), based on [10].

However, as we have not gathered data yet, we rely once again on psychological knowledge to tune the first version of our model. More specifically, in [11], 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 the perceived interpersonal stance of the user, in reaction to a signal or a sentence, and within a certain topic or an entire interaction.

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. In the TARDIS project, enacments of job interviews have been simulated and videos of these interviews have been recorded. To achieve that goal we will annotate these videos with occurences of non-verbal si-

gnals and interpersonal stances ratings, and then use them as the data for learning the parameters of our model.

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

References

[1] M. Argyle. *Bodily Communication*. London : Methuen, 2nd edition, 1988.

[2] A. Arya, L. N. Jefferies, J. T. Enns, and S. 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] N. Bee, S. Franke, and E. Andrea. Relations between facial display, eye gaze and head tilt : Dominance perception variations of virtual agents. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7, 2009.

[5] N. Bee, C. Pollock, E. André, and M. Walker. Bossy or Wimpy : Expressing Social Dominance by Combining Gaze and Linguistic Behaviors. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271. Springer Berlin / Heidelberg, 2010.

[6] E. Bevacqua, E. De Sevin, S. J. Hyniewska, and C. Pelachaud. A listener model : introducing personality traits. *Journal on Multimodal User Interfaces, special issue Interacting ECAs*, 2012.

[7] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions On Computer-Human Interaction*, 12(2) :293–327, June 2005.

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

[9] M. Cavazza, D. Pizzi, F. Charles, and E. Vogt, T.and 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.

[10] A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. 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) :41–48, 2002.

[11] R. Gifford. Mapping nonverbal behavior on the interpersonal circle. *Journal of Personality and Social Psychology*, 61(2) :279–288, 1991.

[12] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *Transactions on Audio, Speech and Language Processing*, 17(3) :501–513, March 2009.

[13] S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth. A conversational agent as museum guide : design and evaluation of a real-world application. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, editors, *Lecture Notes in Computer Science*, pages 329–343, London, UK, 2005. Springer-Verlag.

[14] E. Krumhuber, A. S. R. Manstead, and A. Kappas. Temporal aspects of facial displays in person and expression perception. the effects of smile dynamics, head-tilt and gender. *Journal of Nonverbal Behavior*, 31 :39–56, 2007.

[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] C. Rich and C. L. Sidner. Procedural dialogue authoring with hierarchical task networks and dialogue trees. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents*, IVA'12, 2012.

[17] J. 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.

[18] R. J. Rienks and D. Heylen. Automatic dominance detection in meetings using easily detectable features. In *Workshop on Machine Learning for Multimodal Interaction*, Edinburgh, U.K, 2005.

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

[20] M. Schröder. The SEMAINE API : Towards a Standards-Based Framework for Building Emotion-Oriented Systems. *Advances in Human-Computer Interaction*, 2010 :1–21, 2010.

[21] X. Van Mulken, E. André, and J. Müller. The persona effect : How substantial is it ? In H. Johnson, L. Nigay, and C. Roast, editors, *People and Computers XIII, Proceedings of HCI '98*, pages 53–66. Springer, 1998.

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

[23] J. Wagner, F. Lingenfelser, N. Bee, and E. André. The social signal interpretation framework (SSI) for real time signal processing and recognition. *Proceedings of Interspeech 2011*, 2011.

[24] J. S. Wiggins. *Paradigms of Personality Assessment*. New York : Guilford, 2003.