Leveraging the Dynamics of Non-Verbal Behaviors For Social Attitude Modeling

Soumia Dermouche, Catherine I Pelachaud

Leveraging the Dynamics of Non-Verbal Behaviors For Social Attitude Modeling

Soumia Dermouche and Catherine Pelachaud

Abstract—An Embodied Conversational Agent (ECA) is a virtual character designed to interact with humans in the most natural way. In the recent years, ECAs have been deployed in various contexts, such as commercial consulting and social training. In the context of social training, the virtual agent should be able to express different social attitudes in order to train the user in different situations, likely to occur in real life. Previous studies from psychology underlined the importance of considering the non-verbal behavior as well as its evolution over time, for efficient modeling of interpersonal attitudes. Inspired by these works as well as by advances from sequence mining, we propose to model attitude variation as a sequence of non-verbal signals, each being described by its starting time and duration. We demonstrate the efficiency of our model by integrating the sequences representing attitude variation in an ECA and assessing the obtained results based on the interpersonal circumplex, statistical tests and accuracy measures. To the best of our knowledge, this is the first attempt to study the relationship, in term of perception, between different attitude variations.

Index Terms—Embodied conversational agents, non-verbal behavior, social attitudes, temporal sequence mining.

1 Introduction

In our everyday life, we continuously express different interpersonal attitudes toward each other depending on the interaction context that includes factors such as our interlocutors, our role, our personality, our goal, etc. For example, a person may show dominance in some work contexts while being warm when going out with friends. The same person will not behave in the same way in these different circumstances. She will not display the same behaviors. She may use a more formal language at work, show a more upright posture, smile less, while she may laugh and gesture expressively with friends.

Interpersonal or social attitudes are defined by Scherer as *"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"* [1]. Attitudes are expressed through both verbal and non-verbal behaviors [2]. Moreover, attitudes are *dynamic*: an attitude is *"an affective style"* that *colors* an interaction. Then, attitudes are not only expressed by a given signal at a certain time but rather by the coordination and dynamics of a sequence of multimodal signals whose meaning arises from the interrelation of interactants' behaviors. Several studies underlined the relationship between interpersonal attitudes and several multimodal behaviors such as postures [3], [4], [5], [6], head movement [7], [8], [9], [10] and gaze [11], [12], [4], [13].

Regarding attitude representation, they can be represented by labels like similarity [14] and evaluation [15]. Moreover, attitudes can be represented using a graphic called Interpersonal Circumplex (IPC), where attitudes are plotted along two orthogonal axes: dominance axis (ranging from submissive to dominant) and friendliness axis (ranging from hostile to friendly) [16], [17].

One of the research questions we are addressing is: what makes a person appear more/less dominant or more/less friendly? That is, we are interested in finding out which sequences of non-verbal behaviors trigger a change in the perception of social attitude. The challenge is that the perception of non-verbal signals could be directly influenced by the temporality of signals, namely, order, starting time and duration that could alter their meaning. For example, *averted gaze* is generally related to submission [7], [8], [9], [10], [18]. However, this signal leads to an increase of dominance perception when it is followed by expression of anger [18]. Signal duration is also important for behavior perception. For example, the duration of a *smile* could differentiate between fake and genuine smiles [19], [20]. Keltner demonstrated that the starting time of *smile*, *gaze shift* and *head away* can help differentiating between embarrassment, amusement and shame [21].

Embodied Conversational Agents (ECAs) are virtual characters that can interact autonomously with human users using verbal and non-verbal behaviors [22]. Designing virtual agents have gained a lot of attention in the recent years where virtual agents have become more and more present in our everyday lives. They can be used for a variety of applications ranging from education [23] and training [24] to therapy [25]. We aim to endow an ECA with the ability to express different interpersonal attitudes depending on the interaction context.

To this end, we propose a computational model to endow an ECA with the capability of varying its social attitudes during an interaction. These variations are modeled as sequences of non-verbal behaviors. To encompass the dynamics of non-verbal signals across both modalities and time, we make use of temporal sequence mining. Specifically, we propose a new algorithm for temporal patterns extraction. We implement a fully-automatic method for ex-

- *S. Dermouche is affiliated with Kiliba company, Île-de-France, France. E-mail: soumia.dermouche@hotmail.fr*
- *C. Pelachaud is affiliated with CNRS, laboratory ISIR, Sorbonne Université, Paris, France. E-mail: catherine.pelachaud@upmc.fr*

tracting relevant patterns of non-verbal behaviors that convey variation of social attitudes from a multimodal corpus. We also propose a computational model to generate the agent's non-verbal behavior according to its communicative intentions as well as its attitude variation. Our models are evaluated through perception studies. The novelty of our proposition lies in representing attitude variations as sequences of non-verbal signals, as well as jointly considering the starting time and duration of these signals. In the next section, we give a review of related works in the virtual agent field. Sequence mining task and related algorithms are presented in Section 3. Section 4 describes our methodology for modeling attitude variations as sequences of non-verbal signals. In Section 5, we apply HCApriori algorithm to extract relevant patterns expressing attitude variations. Then, these patterns are simulated into virtual agent and evaluated in Section 6. An attitude planner is developed in Section 7 and evaluated in Section 8. Finally, in Section 9 we conclude and give some perspectives.

2 Related Works

In this Section, we present an overview of the most relevant works related to our topic: attitude modeling for virtual agents. We also focus on the works relying on sequentiality and temporality of non-verbal behavior as key components for human and agent behavior modeling.

2.1 Attitude Modeling for Virtual Agents

Several investigations focused on the impact of some behaviors on the perception of ECA's attitude. Bee *et al.* studied the impact of facial expression, gaze and head direction on the perception of virtual agent's dominance [18]. Later on, this study was completed by adding linguistic behaviors to facial expression, gaze and head direction in order to investigate which modalities contributes the most to the expression of dominance [26]. The authors found that both verbal and the non-verbal channels participate equally to the expression of dominance. Straßmann *et al.* explored the perception of a virtual agent expressing dominance, submission, and cooperativity [27]. Cafaro *et al.* investigated how the interpersonal attitude (hostility/friendliness) and the personality (extraversion) of a virtual agent influences the first impressions of users about the agent [28]. In [29], the authors studied the influence of interruption types (amount of overlap between speakers and utterance completeness) on the perception of interpersonal attitudes during an agent-agent interaction. The results revealed that the interruption types directly influence the perception of attitudes of both agents (interruptee and interrupter): the interruptee is perceived more dominant (and less friendly) when the amount of overlap increases.

Others works focused on the dynamics of attitudes by modeling the evolution of an ECA's attitude over time [30], [31], [32]. The attitudes are first initialized w.r.t. the role of the agent and of its interlocutor. For example, an ECA assuming the role of a policeman will be initialized with a high dominance value when interacting with a gangster and a low dominance value towards his superior. Then, depending on the emotion conveyed by the agent, its attitude is adjusted.

Other research focused on developing computational models for generating agent's behavior according to its attitude. To learn the mapping between attitudes and non-verbal behaviors, a corpus of ECA's non-verbal behaviors conveying attitudes has been gathered and annotated using crowdsourcing [33]. Then, a Bayesian model has been designed in order to automatically generate the non-verbal behavior of the ECA given as input its interpersonal attitude [34]. Using a corpus of job interviews between human recruiter and human job seeker, Chollet *et al.* applied a GSP (Generalized Sequential Pattern) algorithm [35] to extract non-verbal sequences of a recruiter when s/he expresses different interpersonal attitudes toward a candidate [36]. Then, an attitude planner has been developed to generate the behavior of the agent according to its attitude and its communicative intentions. The results showed that most attitudes of the agent were recognized. In this approach, the authors did not consider the temporality (starting time and duration) of non-verbal behavior.

Most of the presented models rely on non-verbal behavior to express a given attitude. However, none leverages the temporal information of these behaviors. Our work addresses this limitation by considering the temporality of the non-verbal behaviors.

2.2 Sequence-Based Multimodal Behavior Modeling

In this Section, we present existing work that encompasses the sequentiality of non-verbal behavior in order to understand and predict phenomena such as emotion and interpersonal attitude.

Niewiadomski *et al.* [37] proposed a constraint-based approach to generate sequences of non-verbal behaviors expressing emotions. Their model includes: (i) a multimodal set of behaviors, extracted from both literature and annotated corpora (e.g., embarrassment is related to ten signals: head down, look down, smile...); (ii) a set of hand-crafted rules to ensure the correct timing and order of behaviors in the sequence. An evaluation study showed that the expressions of emotions with the proposed model are better perceived than when emotions are represented by one signal. However, this approach has some limitations. The corpora the authors used is small and the need for manual work to establish the rules makes the task costly and labor intensive. With and Kaiser used T-Patterns algorithm [38] to detect sequences of facial signals representing five emotions: enjoyment, hostility, embarrassment, surprise, and sadness [39].

Zhao *et al.* used the TITARL algorithm [40] to predict behavioral patterns that convey a variation in interpersonal rapport [41]. For this purpose, a corpus involving a tutor and tutees has been annotated on several levels: gaze, smiles, conversational strategies like social norm violation, and interpersonal rapport. The TITARL algorithm has been applied to extract temporal association rules representing either an increase or a decrease of rapport between tutor and tutees. For example, this patterns *the tutor violates social norms while being gazed at by the tutee, and their speech overlaps within the next minute* characterizes a decrease in interpersonal rapport. The TITARL algorithm has also been used in [42] to extract temporal association rules related to attitude from the SEMAINE database [43]. More precisely,

Janssoone *et al.* investigated the correlation between non-verbal behavior (like eyebrow movements and prosody), and two attitudes: friendliness and hostility. TITARL allows predicting temporal relations between signals such as occurrence interval (e.g., if there is a signal d at time t, then there is a signal c at time $t+5$). However, it does not extract exact duration of signals.

The works in [44], [45], [46] focused on extracting temporal sequences of non-verbal behaviors from human-robot interactions. The extracted sequences have been used to analyze the human's behavior, primarily gaze behavior, in relation to the robot's behavior. In addition to explicitly consider timing of non-verbal behaviors, these works are set in dyadic settings; i.e., they consider both human's and robot's behavior. However, they are not generative; the extracted patterns are not explored for generating robot's behavior.

As we have seen, some works only relied on the order of signals ignoring their temporality [39]. Others works considered a limited number of modalities [44], [39], [45], [47]. Others may rely on hand-crafted constraints [37]. Only a couple of these works explored the extracted sequences of human behaviors for generating a virtual character's behaviors [48]. In our work we are going beyond the limitations of existing works by considering the temporality (starting time and duration) of human behaviors. We propose a fully-automatic, multimodal, sequential and temporal model for extracting non-verbal patterns representing different attitude variations. Afterward, the extracted patterns are used to generate the agent's non-verbal behavior according to its communicative intentions and attitude variation.

3 Temporal Sequence Mining

Sequence mining is a data mining task that aims at discovering relevant patterns hidden in a set of sequences. A pattern is a sub-sequence that occurs frequently in the dataset. Sequence mining has been applied in a wide range of real-life applications in many domains such as marketing [35], bioinformatics [49], text mining [50] and human behavior analysis [24], [44]. In this Section, we present the different algorithms of sequence mining. We introduce a new temporal sequence mining algorithm HCApriori to overcome the limitations of existing algorithms. Finally, we present the metrics that are commonly used to assess pattern quality and that are based on occurrence frequency. We enhance these metrics by considering signal temporality.

3.1 Order-Aware Algorithms

The most widely-used sequence-mining algorithms are Apriori [51], GSP [35] and PrefixSpan [52]. Taking as input a sequence dataset and a given minimum frequency threshold denoted f_{min}, these algorithms discover relevant (frequent) patterns based on signal order. Only patterns (sub-sequences) that occur more than f_{min} are considered as relevant. For example, from the dataset $\{ABB, ABC, CABA, CABCA\}$ with f_{min} = 50% (2 sequences), the frequent patterns of length 3 are $\{CAA, ABA, ABC, CAB\}$. However, relying only on signal's order may become a limitation where timing such as starting time, duration, and delay between signals is informative. For example, for generating the agent's behavior, we need to know when the agent should smile and nod simultaneously and for how long to trigger an increase of friendliness.

3.2 Temporal Sequence Mining Algorithms

Temporal algorithms are designed to address the time-related issues such as: at what moment a temporal signal (cf. Definition 1) happens? And what is its duration? For example, it can extract temporal sequences such as $(A, 2, 6)(B, 3, 8)$. This sequence is interpreted as: the signal A appears from second 2 to second 6 followed by the signal B from second 3 to second 8. These algorithms combine classical sequence-mining algorithms, usually Apriori, with a data clustering algorithm. First, a clustering algorithm allows grouping signals that mostly occur at the same time. The centroid of each cluster will represent one temporal pattern of size one. Then, Apriori-like procedure will be applied repetitively to generate more longer patterns. These algorithms require four inputs: a temporal sequence database; f_{min}; a temporal dissimilarity measure like *CityBlock* used to evaluate the temporal distance between signals; and a dissimilarity threshold (ϵ) used to decide if two signals are temporally similar or not. A temporal sequence is a sequence of temporal signals. A relevant temporal pattern is sub-sequence that occurs at least in f_{min}% of sequences (from the input database) and which has a temporal distance (with these sequences) less than ϵ.

Definition 1. Temporal signal
A temporal signal s is a triplet (t, s, e), where s^t is the signal type (e.g., smile, head nod). s^s and s^e are the starting, respectively the ending, time of the signal (with $s^s < s^e$).

QTempIntMiner [53], QTIPrefixSpan and PESMiner [54] are examples of temporal sequence mining algorithms. These algorithms present two main limitations. First, they do not consider differences of duration between signals. In some cases, signals can have a very different mean duration depending on their type. For instance, in the case of multimodal conversational behaviors, postures and smiles have globally very different mean duration (respectively 32.24 seconds and 2.01 seconds in our corpus). Thus, in such a case, one second represents about 50% of the mean duration of smile and only 3% of posture duration. Table 1 illustrates mean duration of some non-verbal signals. In addition, relying on partitioning clustering algorithms (like, Kmeans), distant signals can be merged into a same cluster which decreases the cluster homogeneity.

3.3 HCApriori Algorithm

To deal with the challenges of the temporal sequence mining algorithms highlighted in the previous section, we propose a new algorithm that we call HCApriori for Hierarchical Clustering Apriori [55]. HCApriori is open source and available on github[1]. The novelty of HCApriori is to **customize the dissimilarity threshold (ϵ)** for each signal type (posture,

1. DOI:10.5281/zenodo.3463304

TABLE 1
Mean duration of some non-verbal signals in seconds.

Signal	Gaze At	Gaze Up	Eyebrow Up	Body Lean	Body Recline	Arms Crossed	Smile
Mean duration	4.46	1.26	2.34	32.34	17.41	11.74	2.01

gesture, gaze, etc.) and to propose an automatic computation of ϵ alternatively to manual settings. For outliers detection, HCApriori relies on **hierarchical clustering** that imposes a distance less than ϵ to the signals from the same cluster. HCApriori operates in two steps: (1) hierarchical clustering is first applied to merge signals into the same cluster if and only if their temporal distance is less than ϵ. At the end of this step, the cluster centroids represent patterns of length one. (2) Taking as input the results of step (1), Apriori-like procedure is adapted to generate longer temporal patterns: based on Apriori algorithm [51], our algorithm generates the frequent temporal patterns in two steps: a set of candidate temporal patterns of length $n+1$ is generated from all the temporal patterns of length n. Then, the infrequent patterns are pruned. Candidate generation and pruning are performed repetitively until no more patterns can be generated.

We evaluate our algorithm on a corpus of job interviews where non-verbal behaviors and attitude variations of the recruiters are annotated. We apply HCApriori to extract patterns of multimodal behaviors characterizing attitude variations. We compare the results of HCAapriori against the results obtained by four state-of-the-art algorithms: QTIPrefixSpan-Kmeans, QTIPrefixSpan-AP, QTIApriori-Kmeans, and PESMiner. The comparison is based on pattern accuracy criteria. The accuracy is defined as the percentage of sequences from the original data that are similar to at least one pattern from the set of extracted patterns[1]. Figure 1 plots the accuracy of the experimented algorithms as a function of f_{min} (minimum frequency threshold). Results showed that HCApriori allows a better extraction of patterns with a significant improvement over the other four state-of-the-art algorithms. The interested reader can find more details about our HCApriori algorithm in [55].

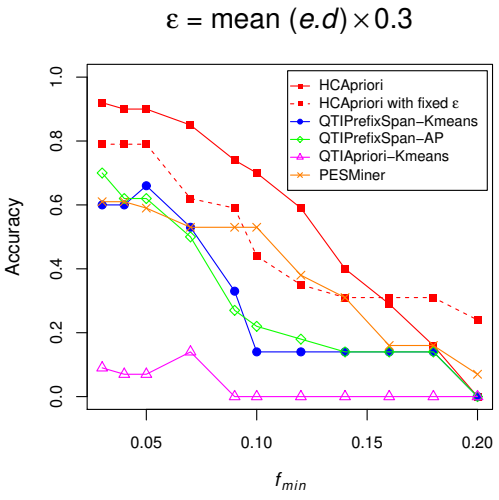


Fig. 1. Accuracies of the compared algorithms for different values of f_{min} ranging from 1% to 20% of the database size. The accuracy value ranges from 0 to 1. HCApriori outperforms the other algorithms and is able to achieve over 0.92 accuracy whereas the runner-up achieves 0.70.

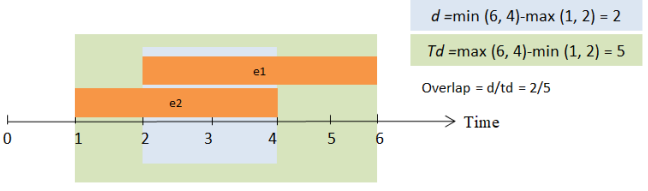


Fig. 2. Representation of overlap between two temporal signals.

3.4 Pattern Quality Measurement

Once the relevant patterns are extracted, we can rank them according to their quality. Based on the occurrence frequency, the quality of an extracted pattern p is assessed using two quality measures: (1) support (Eq. 1) indicates the frequency of the pattern p in the dataset D. D denotes the set of the input temporal sequences. (ii) Confidence (Eq. 2) reflects the proportion of D containing p and expressing the attitude variation v.

$$\text{Sup (p)} = \frac{|\{S \in D : \text{S contains p}\}|}{|D|} \quad (1)$$

$$\text{Conf (p, v)} = \frac{|\{S \in D : \text{S contains p and S expresses v}\}|}{|\{S \in D : \text{S contains p}\}|} \quad (2)$$

In order to provide a temporal similarity between the extracted patterns and the input dataset, we extend the classical measures support and confidence by considering the time overlap between signals. We define the overlap between two temporal signals (cf. Definition 1) e_1 and e_2 in Eq. 3. The overlap represents the duration d ($d = min(e_1^e, e_2^e) - max(e_1^s, e_2^s)$) where two signals e_1 and e_2 appear in the same time windows (in Figure 2, e_1 and e_2 overlap between seconds 2 and 4, then $d = 2$). To normalize the overlap between zero and one, we divide d by the time interval td ($td = max(e_1^e, e_2^e) - min(e_1^s, e_2^s)$) corresponding to the union of e_1 and e_2 (in Figure 2, $td = 5$).

$$\text{Overlap}\,(e_1, e_2) = \begin{cases} 0, & \text{if } d < 0 \\ \frac{d}{td}, & \text{otherwise} \end{cases} \quad (3)$$

The overlap between a pattern p and a sequence S is the sum of the overlap between the signals of p and S. To normalize (between 0 and 1), we divide this sum by the minimum length between p and S.

1. Two temporal sequences are similar if their temporal distance is less than the dissimilarity threshold (ϵ)

The two new measures $SupOverlap$ and $ConfOverlap$, indicating the support and confidence overlap between p and D, are given in Eq. 4 and 5, respectively.

$$\text{SupOverlap (p)} = \frac{\sum_{S \in D} \text{Overlap (p,S)}}{|D|} \quad (4)$$

$$\text{ConfOverlap (p, v)} = \frac{\sum_{S \in D, \text{S expressing v}} \text{Overlap (p,S)}}{|\{S \in D : \text{S contains p}\}|} \quad (5)$$

In the next Section, we describe our methodology for modeling virtual agents displaying variations of social attitude based on HCApriori algorithm.

4 Methodology

Our goal is to develop a virtual agent able to display different attitude variations depending on the interaction context. For example, it can increase its dominance level when training a job candidate passing a job interview. Interpersonal attitudes are conveyed through non-verbal behaviors (e.g., gaze, facial expression, head movements, etc.). Furthermore, attitudes are not only expressed by specific signals but rather by the coordination and dynamics of a series of multimodal signals. In order to encompass both multimodality and dynamic aspects of attitude expression, we represent an attitude variation as a temporal sequence of non-verbal signals. To reach this goal, we follow a five-step methodology:

1) First, we segment a multimodal corpus into four datasets containing non-verbal sequences related to attitude variations (friendliness increase/decrease, dominance increase/decrease).
2) Secondly, we apply HCApriori algorithm to extract for each attitude variation, the relevant patterns expressing this attitude variation.
3) Once the patterns extracted, they are simulated within an ECA and evaluated through a perception study. In this evaluation, the agent speaks non-sense speech as we look only to the perception of social attitude variation through non-verbal behaviors.
4) An attitude planner is developed enabling an ECA to communicate its intentions with variation of social attitudes.
5) Finally, a perception study is conducted to evaluate the attitude planner.

These steps are presented in detail in the following sections.

5 Step 1: Extraction of relevant patterns expressing attitude variations

5.1 Corpus

For attitude variation modeling, we use a corpus of job interviews where a recruiter can express different attitudes toward a candidate [36]. This corpus is composed of three videos showing simulation of three job interviews which involved each time a different middle age human resources practitioner and one youngster candidate looking for a job. We have two men and one woman for both, recruiters and candidates. All participants are French and the job interview were done in French. The goal is to help the candidate for preparing her job interview. The total duration of this corpus is 57 minutes and 32 seconds. The behavior of the recruiters is annotated on two levels: non-verbal behavior and attitude perception. Non-verbal behaviors are annotated by one annotator using the annotation tool Elan [56].

As social attitude varies during the interaction, continuous annotation schema is applied. It allows us to provide a curve of points representing the value of attitude change at every time step of the interaction. We follow the circumplex representation IPC [16] that describes attitudes along two dimensions, namely dominance and friendliness. Using the annotation tool Gtrace [57], the annotation of dominance and of friendliness is done continuously by 12 annotators. In order to avoid content biases from the verbal behavior when annotating attitude dimensions, we filter it out, for both recruiter and candidate, by applying a Pass Hann Band Filter. Each annotator annotates only one dimension of attitude at a time (dominance or friendliness) and the value of annotation ranges from -1 to 1. For dominance dimension (respectively friendliness), -1 means totally submissive (resp. hostile), 0 means "neutral" or "no" dominance (resp. friendliness) and 1 means totally dominant (resp. friendly). The agreement between the annotators in term of Cronbach α is acceptable (α =0.742).

5.2 Non-verbal Behavior Segmentation

We define an attitude variation as:

Definition 2. Attitude variation. *An attitude variation v is a tuple (s, e, t), where v^s, resp. v^e, is the starting, resp. the ending, time of the variation. v^t is the variation type (increase or decrease).*

Having this corpus, we segment the non-verbal behaviors based on attitude variations as indicated in Figure 3. For better annotation quality, we consider the reaction of annotators as recommended in [58] that demonstrated that the accuracy of emotion recognition improves by more than 7% percent when considering the reaction lag of annotators. In SEMAINE corpus the delay varies from one to six seconds [58]. These different values of the delay may come from factors such as the displayed multimodal behaviors, the phenomenon being continuously evaluated or even the annotator's sensitivity. In our study, to choose a value for the reaction lag, we vary its value, lag, from zero to six seconds with a step of 1 second. For each lag value we compute the accuracy of extracted patterns. We find that the $lag = 2$ second gives the best accuracy result. Thus, we choose this value for the reaction lag.

For each variation v occurring when the recruiters are speaking, we collect all non-verbal signals that appear during this variation (that is, between $v^s - lag$ and $v^e - lag$). These signals compose a sequence S in which the starting time, respectively, the ending time of each non-verbal signal s is the time difference between the starting time, respectively, the ending time of s and v^s minus lag. For example, in Figure 3, the second variation of dominance increase starts at 6 seconds and finishes at 15 seconds after taking out the reaction lag. Two signals appear during this variation: *head shake* (from seconds 6 to 10) and *arms crossed* (from seconds 9 to 17). Then, the temporal sequence representing this variation is (*head shake*, 0, 4) followed by (*arms crossed*, 3, 11). This segmentation allows us to build four sets

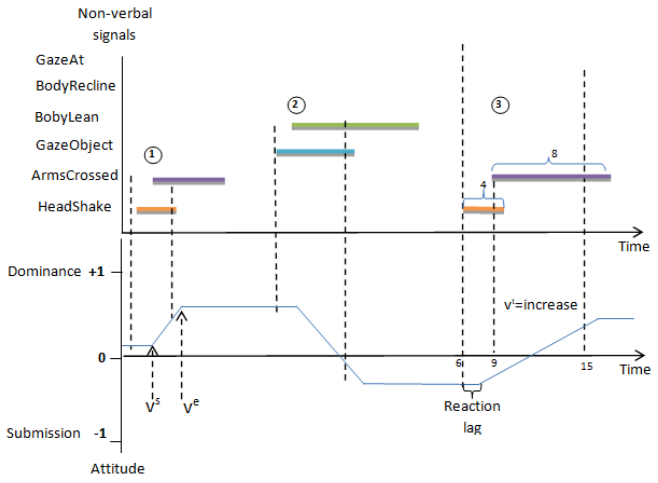


Fig. 3. Non-verbal behavior segmentation based on attitude variations. The result is a set of non-verbal sequences for each type of attitude variation.

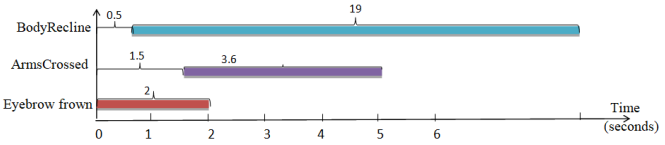


Fig. 4. An example of a pattern representing dominance increase. This pattern can be interpreted as follows: the recruiter frowns his eyebrows for 2 seconds. After 1.5 seconds, he crosses his arms for 3.6 seconds while leaning backward.

of non-verbal behavior sequences representing four types of attitude variation: dominance increase (143 sequences), dominance decrease (110 sequences), friendliness increase (80 sequences) and friendliness decrease (94 sequences).

We also extract sequences of behaviors occurring when "no" attitude is expressed. We define by "no" attitude expression, the segments of the corpus that are marked with an attitude value around zero (for commodity, we take all values between -0.05 and 0.05). We refer to these extracted sequences as "reference" with "no" attitude expression. Then, since annotators consider one dimension at a time, we obtain two datasets of sequences representing respectively "dominance reference" (40 sequences) and "friendliness reference" (30 sequences).

5.3 Step 2: Pattern Extraction

After the segmentation of the corpus, we apply HCApriori algorithm (described in Section 3.3) to extract patterns related to each attitude variation. To choose the values of f_{min} and ϵ we rely on pattern accuracy. We choose the values that give an accuracy superior than 70% (see Figure 1). So, we apply HCApriori by fixing f_{min} to 10% of the dataset size and ϵ to 30% of the mean duration of each signal type. We obtain, respectively, 165, 262, 210 and 156 patterns for, respectively, dominance increase, dominance decrease, friendliness increase and friendliness decrease. An example of an extracted pattern representing dominance increase is illustrated in Figure 4.

6 Step 3: Evaluation of the Extracted Patterns

In order to evaluate which attitude variations are conveyed by the extracted patterns, we design a perception experiment where a virtual agent displays these patterns.

6.1 Experimental Design

We evaluate four different categories of non-verbal patterns denoting four attitude variations: dominance increase (*DomInc*), dominance decrease (*DomDec*), friendliness increase (*FrInc*) and friendliness decrease (*FrDec*). For each of them, we evaluate four non-verbal patterns. We also rate two patterns for the two "reference" attitudes. Using the virtual agent platform called GRETA-VIB [59], we generate videos showing an agent displaying these patterns. Examples of generated videos are available here: https://youtu.be/ouiwShHfe3I. Since we are interested in modeling an agent that talks with different attitudes, we associate unintelligible speech with these sequences of non-verbal behaviors. We have left aside the content of speech which might also contribute in the perception of attitudes [60]. We generate unintelligible speech by giving Arabic text as input to an English text-to-speech synthesizer. We produce a total number of 18 videos: 16 **comparison videos** (4 attitude variations $\times$ 4 patterns) and two **reference videos**: "dominance reference" (denoted *DomRef*) and "friendliness reference" (denoted *FrRef*).

The evaluation follows a two-step process: first participants are asked to view and rate the ECA in the reference video (*DomRef* for the conditions *DomInc* and *DomDec* and *FrRef* for the conditions *FrInc* and *FrDec*). Then, participants view four pairs of videos where each pair is made of the reference video and a comparison video; they are asked to rate the behavior of the ECA in the comparison video. Participants are randomly assigned to one condition in which the ECA displays patterns expressing one given attitude variation. Videos appear automatically once participants view the whole current video and answer all questions. The order of videos is shown according to a latin square design to control first-order carryover effects [61].

6.2 Measures

Participants evaluate their perception of agent's attitude along several adjectives. To find the most relevant adjectives that characterize the perception of attitude, we use interpersonal circumplex (IPC) measurements. Graphically, the IPC is represented by two orthogonal axes: a vertical axis for dominance and a horizontal axis for friendliness [16], [17] (cf. Figure 5). Thus, each interpersonal disposition (like "forceful") can be represented, within the IPC, as a weighted combination of dominance and friendliness.

Most IPC measurements split the IPC into eight octants or scales that are alphabetically labeled counterclockwise: PA, BC, DE, FG, HI, JK, LM and NO (cf. Figure 5). Each octant is represented by several adjectives; e.g., assured and dominant are in the PA octant. Locke provided an overview of IPC measurements [62]. The Interpersonal Check List (ICL) [16] and the Interpersonal Adjective Scales (IAS) [63] are two measures for rating interpersonal traits.

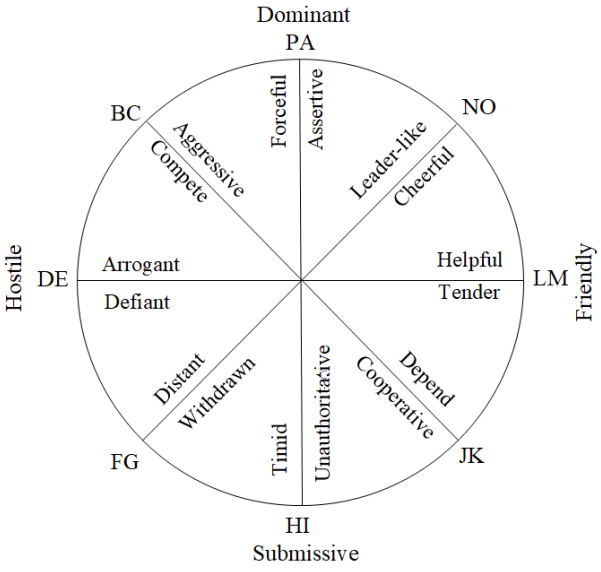


Fig. 5. Interpersonal circumplex (IPC) [16]. The IPC is represented by two orthogonal axes dominance and friendliness and it is splitted into eight octants. Each octant is characterized by a set of adjectives.

To measure the perception of attitudes, previous researches relied on either IAS [36], [64], [29], [42], [32] or ICL [65]. In our work, we choose to use a combination of both IAS and ICL.

For simplifying the rating, two adjectives, one with the highest factor in IPC and one in IAS, are selected from the analysis done respectively in [16] for ICL and in [66] for IAS. In total, we use 16 adjectives: 8 adjectives from IAS (assertive, aggressive, arrogant, distant, timid, cooperative, tender and cheerful) and 8 from ICL (forceful, compete, defiant, withdrawn, unauthoritative, depend, helpful and leader-like) (see Figure 5).

Unlike the previous studies where participants only rated the perception of one attitude dimension at a time [36], [64], [29], [42], [32], we ask the participants to rate the perception of the two dimensions simultaneously to discover the relationship, such as halo and compensation effects, that may exist between the two dimensions. Compensation effect is a negative relationship between two dimensions of social judgment [67]. Halo effect on the other hand is a positive relationship between two dimensions: changing one dimension involves a change in the other dimension in the same direction [67]. For this purpose, participants rate the behavior of the agent by answering 16 questions related to the 16 selected adjectives: *in your opinion, is the behavior of the virtual character assertive?*. All answers are on a 5-point labeled Likert scale (1 =“strongly disagree”, 2 = “partially disagree”, 3 = “neutral”, 4 = “partially agree”, and 5 = “strongly agree”). We add an attention check strategy in order to detect and filter out the participants who randomly responded to questions. For this, we ask a trap question about the color of the agent’s hair (the ECA used in this experiment is blond). The order of this question is shown according to a latin square design.

6.3 Hypotheses

Our hypotheses are:

- **H.Ref**: for *DomRef* and for *FrRef*, the ECA will be evaluated as expressing “no” attitude.
- **H.Dom**: for *DomInc*, the ECA will be evaluated as **more dominant** compared to the ECA in *DomRef*.
- **H.Sub**: for *DomDec*, the ECA will be evaluated as **more submissive** compared to the ECA in *DomRef*.
- **H.Fr**: for *FrInc*, the ECA will be evaluated as **more friendly** compared to the ECA in *FrRef*.
- **H.Hos**: for *FrDec*, the ECA will be perceived as **more hostile** compared to the ECA in *FrRef*.

6.4 Results

We recruit a total of 64 participants via CrowdFlower, 42% of them are between 21 and 30 years old, 33% between 31 and 40, 21% between 41 and 50, 4% are more than 50 years old, 85% are male, 53% have a master level, 57% are Spanish, 15% are French and 15% are German. We analyze the results in three different ways by: (1) plotting the results on the IPC, (2) investigating significance of the results and (3) computing the recognition rate of attitude variations.

6.4.1 Measurement scoring: circular profile

To score the results, we follow the procedure described in [68]. This procedure can be used to score data from any IPC inventory. It is composed of three steps:

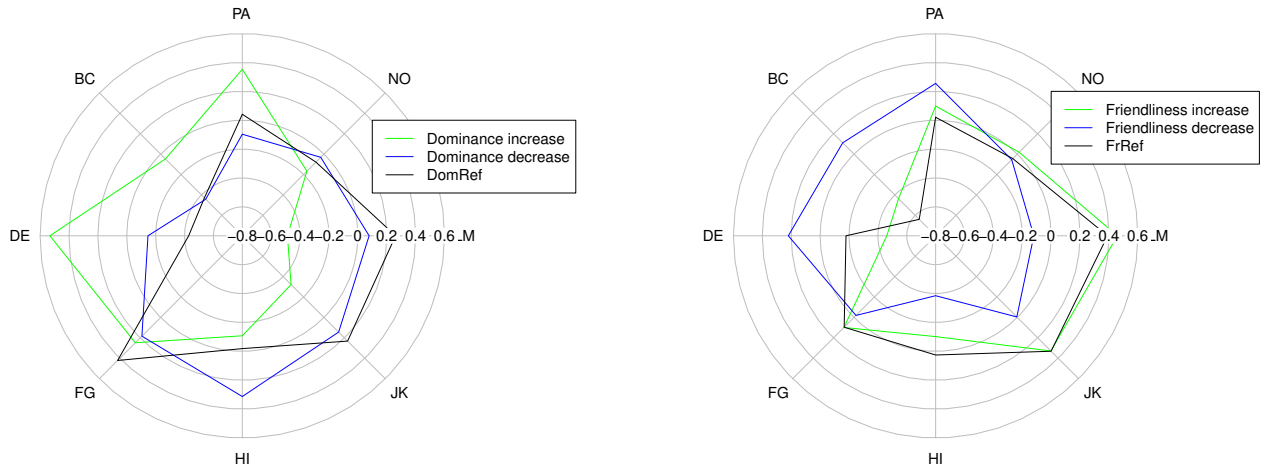
1) Compute the general factor score by averaging the eight octant scores.
2) Ipsatize octant scores by subtracting the general factor score from each octant score.
3) Plot the ipsatized scores on the IPC ranging from the lowest value to the highest value.

Figure 6 plots the ipsatized scores for each condition. We can observe that: (i) for *DomInc* and *FrDec*, the ECA is perceived as: **more dominant (PA)**, **more hostile (DE)** and **less friendly (LM)** compared to the ECA in, respectively, *DomRef* and *FrRef*. (ii) For *DomDec*, the agent is evaluated as **more submissive (HI)** compared to the ECA in *DomRef*. (iii) The ECA in *FrInc* is perceived as **equivalent** to the ECA in *FrRef*.

6.4.2 Result significance

By plotting the agent profile on the IPC (see Figure 6), we can visually interpret how the agent is perceived by participants. We also perform statistical tests to investigate if these results are statistically significant or not. As our data is not normally distributed (Shapiro test’s $p < 0.5$), we use paired Wilcoxon test (equivalent to t-test) to check if there are significant differences in the perception of the agent between the reference video and the comparison video. The revealed differences are summarized as follows:

1) ECA in *DomInc* is evaluated as **more dominant** (*aggressive* ($V = 0$, $p < .005$) and *forceful* ($V = 8$, $p < .05$)) compared to the agent in *DomRef*, therefore **H.Dom** is supported. The agent is also perceived as **more hostile** (*compete* ($V = 6$, $p =< .05$), *arrogant*



Dominance (a) **Friendliness (b)**

Fig. 6. Circular profile of the agent when displaying the four attitude variations and the two reference attitudes. For *DomInc* and *FrDec*, the ECA is perceived as more dominant, more hostile and less friendly compared to the ECA in the reference videos. The agent conveying dominance increase is evaluated as more submissive compared to agent in *DomRef*.

($V = 3.5$, $p < .005$), *defiant* ($V = 0$, $p < .005$), and *distant* ($V = 8.5$, $p < .05$)) and **less friendly** (*cheerful* ($V = 86$, $p < .005$), *helpful* ($V = 88$, $p < .005$), *cooperative* ($V = 103.5$, $p < .005$), and *tender* ($V = 97.5$, $p < .05$)) compared to the agent in *DomRef*. We observe that increasing dominance influences not only the perception of dominance but also the perception of friendliness. These results highlight a compensation effect between the perception of dominance and of friendliness: increasing dominance leads to a perception of friendliness decrease and hostility increase.

2) For *DomDec*, the ECA is evaluated as **more submissive** (*timid* ($V = 14$, $p < .05$) and *unauthoritative* ($V = 21.5$, $p < .05$)) compared to the agent in *DomRef*, thus the hypothesis **H.Sub** is accepted. The agent is also perceived as **more friendly** (*cheerful* ($V = 22$, $p < .05$)) compared to the agent in *DomRef*. These results underline another compensation effect between the two attitude dimensions: decreasing dominance leads to friendliness increase.
3) For *FrDec*, the ECA is perceived as **more hostile** (*arrogant* ($V = 18$, $p < .05$)) compared to the agent in *FrRef*, therefore **H.Hos** is validated. In addition, the agent is evaluated as **more dominant** (*aggressive* ($V = 0$, $p < .005$) and *forceful* ($V = 2$, $p < .005$)) compared to *FrRef*. Thus, we find another compensation effect: decreasing friendliness leads to dominance increase.
4) For *FrInc*, participants rated the ECA as **equivalent** to the ECA in *FrRef*, therefore **H.FR** is rejected.

6.4.3 Comparison of patterns within the conditions

We want to understand if a given pattern, from the four patterns (p_1, p_2, p_3, p_4), we use to evaluate the attitude variations, has an impact in the perception of an attitude change. Table 2 gives the four patterns used for the condition *DomInc*. To explore the effects of the four patterns within their respective four attitude variations, we conduct a Friedman test as our data is not normally distributed. Friedman test is a non-parametric test alternative to the one-way ANOVA with repeated measures. No significant differences between the four patterns are detected for friendliness increase, friendliness decrease and dominance decrease.

For dominance increase, results reveal a significant difference between the four patterns characterizing this attitude change. This difference concerns the evaluation of three adjectives: *compete* ($F(3) = 13.9$, $p < .005$), *timid* ($F(3) = 18.21$, $p < .001$), and *unauthoritative* ($F(3) = 14.78$, $p < .005$). Bonferroni post-hoc test shows that the ECA displaying pattern P_3 is evaluated significantly more *timid* ($p < .05$) and *unauthoritative* ($p < .05$) than the ECA displaying P_2 and P_4. Moreover, the agent displaying P_4 is evaluated as more *compete* ($p < .05$) compared to the agent showing P_3.

6.4.4 Recognition accuracy of attitude variations

In order to assess how accurate is the recognition of the attitude variations, we cast the problem as multi-label classification task, where the predicted class (label) can be one or more among the four attitude variations (*DomInc*, *DomDec*, *FrInc*, and *FrDec*). Thus, we use classical measures from Information Retrieval: recall, precision and F-measure. The recall of a given adjective A represents the number of videos representing A (e.g., the videos related to *DomInc* represent the two adjectives forceful and assertive) and evaluated as expressing A relative to the total number of videos. For each attitude variation, the total number of videos is 64 videos resulting from 16 participants $\times$ 4 evaluated patterns. Each attitude variation has two representative adjectives: forceful and assertive for *DomInc*, unauthoritative and timid for *DomDec*, helpful and tender for *FrInc*, and finally defiant and arrogant for *FrDec*. We consider that a given video is evaluated as expressing A if the participant's response for A is either "partially agree" or "totally agree". The precision of a given adjective A is defined as the number

TABLE 2
The four evaluated patterns for *DomInc*.

P_1:	(*Body Recline*, 1, 20) (*Eyebrow down*, 2, 4) (*Arms Crossed*, 10, 20) (*Beat*, 9.25, 10.45)
P_2:	(*Beat*, 0.65, 2.65) (*Head shake*, 2.15, 4.26) (*Beat*, 5.5, 7.5) (*Eyebrow up*, 6.1, 8.2)
P_3:	(*Beat*, 1.1, 3.2) (*Beat*, 5, 6.8) (*Arms crossed*, 7, 20) (*Eyebrow down*, 9.6, 10.9)
P_4:	(*Arms crossed*, 0.65, 20.85) (*Beat*, 0.65, 2.8) (*Head shake*, 2.15, 4.15) (*Head side*, 5.45, 9.4)

of videos representing A and evaluated as expressing A relative to the total number of videos evaluated as expressing A. For example, 31 videos representing dominance increase are assigned to *forceful* then the recall of *forceful* is 48.43% (31/64). Also, 6 videos representing dominance decrease, 7 videos representing friendliness increase, and 48 representing friendliness decrease are rated as expressing the adjective *forceful*. Consequently, the precision of *forceful* is 33.69% (31/(31+6+7+48)). F-measure is finally computed as the harmonic mean of recall and precision. Each measure is calculated for each attitude variation (condition) by averaging the results obtained from its representative adjectives.

TABLE 3
Recall, precision, and F-measure for each attitude variation.

	DomInc	DomDec	FrInc	FrDec
Recall	39%	40%	35%	35%
Precision	34%	43%	36%	31%
F-measure	36%	41%	35%	33%

As we can see in Table 3, the best results are achieved for *DomDec*. The recall for the four conditions is less than 50% which means that only less than half of videos expressing a given attitude variation are recognized by participants as expressing this attitude variation. The precision is less than 50% for all attitude variations, which means that, for each attitude variation, more than half of the videos evaluated as expressing this variation are actually assigned to another attitude variation. In Table 4, we report the distribution of the predictions over the actual conditions of the predictions. A cell in this Table (where actual=A and predicted =B) gives the number of videos actually expressing A and evaluated as expressing B. From these results, we validate the compensation effects given in Section 6.4.2. In addition, we observe that for both reference videos, participants perceive the agent to be **friendly** but not **hostile**, nor **submissive**.

TABLE 4
Distribution of the predictions over the actual conditions. The predictions highlighting the compensation effects given in Section 6.4.2 and the friendliness perception of the agent in *DomRef* and *FrRef*.

		Predicted			
		DomInc	*DomDec*	*FrInc*	*FrDec*
Actual	*DomInc*	**39%**	22%	20%	**36%**
	DomDec	25%	**40%**	**40%**	15%
	FrInc	13%	22%	**35%**	26%
	FrDec	**51%**	18%	7%	**35%**
	DomRef	34%	18%	**52%**	4%
	FrRef	20%	12%	**43%**	10%

6.5 Discussion

The reference videos are generated from the non-verbal sequences that were characterized with attitude values close to zero. We assume that the agent in these videos would be perceived as expressing "no" attitude. To our surprise, the result of the study shows that the agent is evaluated as friendly which invalidates the hypothesis **H.Ref**. We find no significant difference in the perception of the agent in the *FrRef* and in the *FrInc* condition. The hypothesis **H.Fr** is not validated. An explanation could be that, since the agent in the reference videos is already evaluated as friendly, the agent in the *FrInc* is not perceived as being significantly more friendly than the agent in *FrRef*. The three other hypotheses, **H.Dom**, **H.Sub** and **H.Hos**, are validated.

For *DomInc*, we find a main effect of the four evaluated patterns on the perception of the ECA's attitude. The revealed differences concern 3 out of the 16 adjectives. These differences can be caused by any parameters defining the pattern of behaviors. A more thorough study needs to be conducted to understand this. In order to understand the impact of each pattern on the perception of dominance increase, we redo the statistical test four times, considering one pattern at a time. The same results have been obtained as when considering the four patterns all together. We conclude that, even-though the four patterns show significant differences along three adjectives, all four patterns are perceived as conveying a dominance increase.

According to the representation of attitudes on the interpersonal circumplex, the two poles of an attitude dimension (dominance/submission, friendliness/hostility) are symmetrical with respect to the center of the circumplex. As a result, it is expected that the increase of an attitude toward a given pole would result in a decrease in the perception of the opposite pole. For example, an increase of friendliness would decrease the perception of hostility and vice versa. Based on the circular profile (see Figure 6), for both poles of each attitude dimension, this relationship is observed in both directions of the attitude variations.

Several works on attitude modeling rely on the assumption that there is a compensation effect between the two attitude dimensions. To compute which social attitudes an agent conveys to its interlocutor, the authors in [30], [**?**] defined a set of rules such as positive emotions conveyed by the agent increase its friendliness and decrease its dominance toward the user. Vice versa, negative emotions decrease its friendliness and increase its dominance. Other works rely on the interpersonal complementary theory [16], [17] to model the attitude of agents [64]. According to this theory, two persons should express complementary or anti-complementary attitudes in order to maintain an interaction: expressing similar attitudes on the friendliness dimension and opposite attitudes on the dominance dimension. But, to the best of our knowledge, there are no studies, in term of perception, on the interrelation between attitude dimensions. To better understand this interrelation, we evaluate

both attitude dimensions at the same time. As a consequence, we underline a compensation effect between the perception of dominance and of friendliness drawn from the following observations:

- the perception of dominance increase leads to the perception of friendliness decrease.
- the perception of dominance decrease leads to the perception of friendliness increase.
- the perception of friendliness decrease leads to a perception of dominance increase.

We find high correlation between the perception of dominance increase (*DomInc*) and the perception of friendliness decrease (*FrDec*). A possible explanation is that some non-verbal signals have the same effect on the perception of dominance and of hostility [69], [70], [3], [33]. For example, both dominance and hostility are characterized by a negative facial expression and no gaze avoidance [69], [70], [3], [33].

To sum up, three out of the five hypotheses (**H.Dom**, **H.Sub** and **H.Hos**) are validated. So the sequences expressing the corresponding attitude variations are properly recognized. This supports our assumption that attitude variations can be represented as sequences of temporally-ordered non-verbal signals. The next step is to use the extracted patterns to build an attitude planner that computes the non-verbal behaviors for Embodied Conversational Agents.

7 Step 4: Sequential Attitude Planner Model

Step two of our methodology extracts patterns of non-verbal signals conveying attitude variations (see Section 5). These patterns are extracted regardless of the verbal content. In order to have a virtual agent communicating with a variation of social attitude, our main idea is to combine the non-verbal signals of pattern expressing the agent's attitude variation with the behaviors communicating its other intentions. For that, we integrate our computational model of social attitude variation into a virtual agent platform GRETA-VIB [59].

GRETA-VIB follows the SAIBA framework composed of three main modules, namely the intent planner, the behavior planner and the behavior realizer [71]. The intent planner computes communicative intentions the agent aims to communicate. The behavior planner instantiates these intentions into multimodal behaviors. For example, the communicative intention *greet* can be expressed by either, a *hand gesture*, a *facial expression (smile)* or *an eyebrow movement*. Finally, the behavior realizer computes the corresponding animations. To model social attitudes within the agent platform, we include in the behavior planner a new module called *Sequential Attitude Planner*. The next sections go through the details of the new agent framework.

7.1 Intention Sequence Generation

The communicative intent planner generates a sequence of non-verbal behaviors expressing the communicative intentions specified with the Functional Markup Language (FML) [72]. Once all communicative intentions are instantiated by the behavior planner, we obtain a sequence of multimodal behaviors that we call *intention sequence* (S_{int}). In the example described in Figure 7.1, the FML contains three intentions: *emphasis*, *performative (ask question)*, and *question marker (boundary tone)* associated to the sequence S_{int}: *head*, *gesture*, *face* and *gaze*.

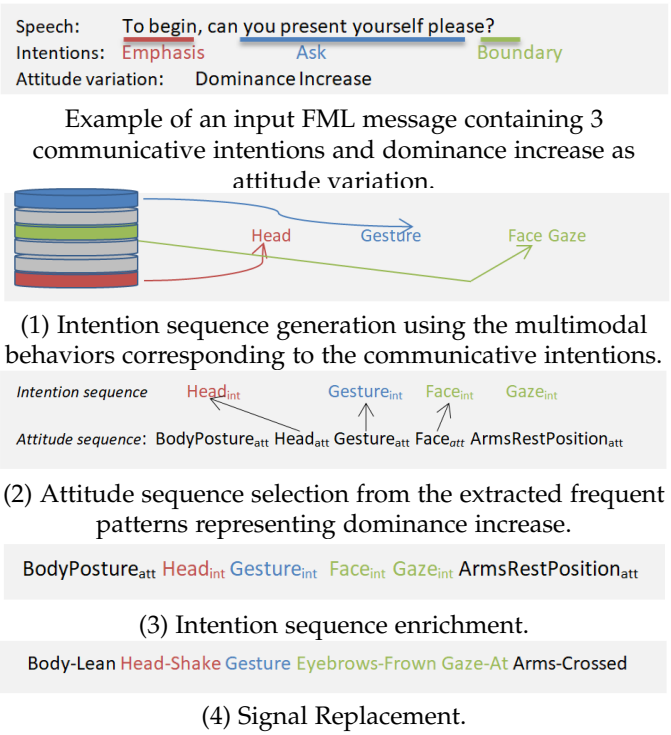


Fig. 7. Outline of the sequential attitude planner model.

7.2 Attitude Sequence Selection

Once the communicative intentions are instantiated, the next step is to choose, from the extracted patterns (cf. Section 5), the most appropriate pattern that we call *attitude sequence* (S_{att}) conveying the desired attitude variation (V). Appropriateness of S_{att} is defined here as the most representative sequence for conveying the attitude variation V and as the most similar to S_{int}. The representativity of S_{att} for expressing the attitude variation V is a combination of support (Eq. 1 and Eq. 4) and confidence (Eq. 2 and Eq. 5) as indicated in Eq. 6.

$$\begin{aligned} AttitudeRep(S_{att}, V) = Conf(S_{att}, V) \times Sup(S_{att}) \\ \times ConfOverlap(S_{att}, V) \times SupOverlap(S_{att}) \end{aligned} \quad (6)$$

The similarity between S_{int} and S_{att} is evaluated in terms of the presence of multimodal behaviors and of their temporality (*overlap*) as defined in Eq. 7. $SimType$ returns the number of behaviors from sequences S_{int} and S_{att} that are of the same modality while $Overlap(S_{int}, S_{att})$ represents the time overlap between S_{int} and S_{att}. We consider the non-verbal modalities: *gesture*, *gaze*, *head*, *posture*, *arms rest position* and *facial expression*. In the example of the Figure 7.2, $SimilarityType = 3$, we have three non-verbal modalities (*head*, *gesture*, and *facial expression*) that are present in both S_{int} and S_{att}.

$$Similarity(S_{int}, S_{att}) = SimType(S_{int}, S_{att}) \times Overlap(S_{int}, S_{att}) \quad (7)$$

In this step, we also associate behaviors from S_{int} to their mapped behaviors from S_{att}: $head_{int} \rightarrow head_{att}$, $gesture_{int} \rightarrow gesture_{att}$, and $face_{int} \rightarrow face_{att}$.

7.3 Intention Sequence Enrichment

In the next step, we enrich the set of communicative behaviors with the set of attitude behaviors. It is obtained by merging both S_{int} and S_{att}: each behavior b_{att} in S_{att} that does not appear in S_{int} is added to S_{int}. Using the same example from Figure 7.3, we add the behaviors *body posture* and *arms rest position* to S_{int}. The timing of b_{att} remains the same if it does not overlap with another signal b_{int} in S_{int} from the same modality of b_{att}. Otherwise, we adjust the timing of b_{att} to allow the agent to display the behavior b_{int}. Time adjustment of b_{att} is indicated in Eq. 8. For example, if the agent has the intention to ask a question instantiated by a gesture (b_{int}) from 0 to 1.5 sec. and, in the S_{att}, we have another gesture (b_{att}) from 1 to 4 sec. then, the agent will play the gesture b_{att} when it finishes doing b_{int} (at 1.5 sec. instead of 1 sec.).

$$\begin{cases} b^e_{att} = b^s_{int} & \text{if overlap } (b_{att}, b_{int}) > 0 \text{ and } b^e_{att} > b^s_{int} \\ b^s_{att} = b^e_{int} & \text{if overlap } (b_{att}, b_{int}) > 0 \text{ and } b^e_{int} > b^s_{att} \\ \text{No adjustment,} & \text{otherwise} \end{cases} \quad (8)$$

7.4 Signal Replacement

In order to represent the relationship between non-verbal behaviors and attitude variations, we compute the frequency of occurrence of a given behavior b with respect to a given attitude variation V. We consider that a behavior b_1 is more representative of an attitude variation V than a behavior b_2 if the frequency of occurrence of b_1 is higher than the frequency of occurrence of b_2.

Finally, our model will replace each behavior b_{int} in the S_{int} with its mapped behavior b_{att} in the S_{att} if the frequency of b_{att} is higher than the frequency of b_{int}. In the example from Figure 7.4, the attitude planner chooses b_{att} (*Eyebrows-Frown*) as final signal because the frequency of this signal for dominance increase is higher than the frequency of $face_{int}$.

8 Step 5: Evaluation of the Sequential Attitude Planner Model

To evaluate the Sequential Attitude Planner, we conduct a perception study where we measure how a virtual agent is perceived when communicating its intentions with different attitude variations. Since we used a job interview corpus to extract relevant non-verbal patterns related to different variations of the recruiter's social attitudes, we keep a similar scenario for this evaluation study. The ECA plays the role of recruiter interviewing for a job opening.

8.1 Experimental Design

We design an empirical experiment in which participants compare a set of video pairs. Each pair is made up of a video of the virtual recruiter with "no" attitude variation and a video with an attitude variation. In each video the agent says a sentence with a given attitude variation. We choose seven sentences (questions) that have a rather "neutral" verbal content. An example of sentence is: *if we decided to offer you this job, when would you be ready to start?*. Seven reference videos (*ref*) are generated without our *sequential attitude planner* (i.e. displaying no attitude change) and 28 reference videos are generated with our *sequential attitude planner* (4 attitude variations $\times$ 7 sentences). Examples of generated videos are available here: https://youtu.be/SursS3oXavk.

We evaluate five experimental conditions corresponding to the four attitude variations: dominance increase (*DomInc*), dominance decrease (*DomDec*), friendliness increase (*FrInc*), friendliness decrease (*FrDec*) as well as reference attitude (*Ref*). The five conditions are tested in a between-subjects design.

Participants are assigned to one condition. If the condition is *Ref* then participants are asked to view seven reference videos and rate each of them by answering 20 questions such as *"in your opinion, the behavior of the virtual character is assertive?*. For the other conditions (*DomInc, DomDec, FrInc, FrDec*), participants view and compare seven pairs of videos: reference video vs. comparison video and rate the behavior of ECA in the comparison video. An example of comparison question is: "*compared to the reference video, is the behavior of the virtual character in the comparison video dominant?*. In addition to the 16 adjectives used in the first experiment, we add four new adjectives: *dominant*, *submissive*, *friendly*, and *hostile*. All answers are on a 5-point labeled Likert scale (1 ="strongly disagree", 2 = "partially disagree", 3 = "neutral", 4 = "partially agree", and 5 = "strongly agree"). The synthesized speech is identical for each pair of videos. We are aware that voice may also reveal a change in attitude [42]. But, since, on the one hand we have not focused our work on the acoustic feature of attitude change, and on the other hand, most speech synthesizers do not model attitude change, we decide to use neutral voice for each video of each condition. After viewing the current pair of videos and answering the questions related to the agent's behavior, another pair of videos is automatically displayed. We show questions and videos according to a latin square design. We rely on the same hypotheses as the first experiment (cf. Section6.3).

8.2 Results

We have a total of 91 participants contacted via CrowdFlower (18 for each condition, one is excluded based on the attention check question), only 20% are female, 50% had a master level and all participants are Europeans or Americans (34% Spanish, 20% German and 18% French). 39% of them are between 21 and 30 years old, 36% between 31 and 40, 19% between 41 and 50, 6% are over 50 years old.

To investigate the effects of the seven videos within the five conditions, we conduct a Friedman test. We choose this test because our data is not normally distributed. For the five conditions, we do not find any significant effect of videos on attitude perception. Then, to assess whether a significant difference exists between the perception of the reference and of the comparison videos, we conduct an unpaired Wilcoxon test. For *FrInc* and *DomDec* no significant differences are detected. For *DomInc*, the agent is perceived as:

- **More dominant**: *aggressive* ($W = 183$, $p < .001$), *forceful* ($W = 165.5$, $p < .005$) and *dominant* ($W = 161$, $p < .005$) compared to the agent in *ref*, therefore **H.Dom** is supported;
- **More hostile**: *arrogant* ($W = 182.5$, $p < .001$), *defiant* ($W = 176.5$, $p < .001$), *distant* ($W = 160.5$, $p < .05$) *hostile* ($W = 182.5$, $p < .001$) compared to the agent in *ref*;
- **Less friendly**: *helpful* ($W = 19.5$, $p < .001$), *cheerful* ($W = 36.5$, $p < .005$), *tender* ($W = 40.5$, $p < .05$), and *friendly* ($W = 0$, $p < .001$) compared to the agent in *ref*.

Finally, for *FrDec*, the ECA is evaluated as:

- **More dominant**: *aggressive* ($W = 163.5$, $p < .005$);
- **More hostile**: *arrogant* ($W = 151$, $p < .05$), *hostile* ($W = 157$, $p < .05$), thus **H.Hos** is accepted;
- **Less friendly**: *cooperative* ($W = 36$, $p < .005$), *helpful* ($W = 28.5$, $p < .001$), and *friendly* ($W = 23.5$, $p < .005$).

8.3 Discussion

The two hypotheses **H.Dom** and **H.Hos** are supported. As in the first evaluation study, the agent in the reference condition is perceived as friendly, thus the hypothesis **H.Ref** is rejected. The agent expressing friendliness increase is evaluated as equivalent to the agent in the reference video, therefore, **H.Fr** is rejected.

Unlike our first study, the hypothesis **H.sub** hypothesis is not validated. Chollet and colleagues [36] found similar result with their virtual recruiter displaying dominance decrease. This result can be related to the context of the interaction where the agent plays the role of a job recruiter. In such context, the recruiter tends to control the interaction and therefore appears naturally dominant and not submissive. This change in perception confirms the importance of the interaction context that can alter the perception of attitude.

An attitude dimension is represented by two symmetrical poles (dominance/submission, friendliness/hostility). We are expecting a negative relationship between the two poles of an attitude dimension (an increase of a given pole would result in a decrease in the perception of the opposite pole). This assumption is statistically significant for friendliness/hostility: when the agent is evaluated as more hostile, it is also perceived as less friendly, and vice-versa.

We can conclude from this study that our attitude planner allows the ECA to express a variation of attitudes, in particular *dominance increase* and *friendliness decrease*. The non recognition of the variation *dominance decrease* could be caused by the interaction context, here the role of the agent. The friendliness perception of the agent in the reference condition seems to affect the recognition of the variation *friendliness increase*.

9 Conclusions

In this work, we described a computational model for extracting and generating sequences of non-verbal behaviors representing attitude variations. The originality of our model is to explicitly consider the temporality of non-verbal signals which is achieved using our temporal sequence mining algorithm HCApriori. The extracted sequences are integrated into the GRETA-VIB platform to allow an ECA to express different attitude variations. We evaluated the non-verbal sequences extracted with HCApriori algorithm and the multimodal sequences generated with our sequential attitude planner. Our results revealed the relationship, in terms of perception, between different attitude variations. In particular, we found a high correlation between the perception of dominance increase (dominance) and friendliness decrease (hostility). A possible explanation is that both attitudes, dominance and hostility, can be expressed by the same non-verbal behaviors [69], [70], [3], [33]. This assumption needs to be more thoroughly analyzed and validated. Finally, the results confirmed the importance of the interaction context in the perception of multimodal behaviors: in a context-free scenario (as in the first evaluation), the agent expressing a decrease in dominance is perceived as such; but when the agent is in a given scenario (e.g., acting as a job recruiter (second evaluation)) the agent is no longer perceived as conveying this attitude variation.

Despite the achieved results, our work still present some limitations. Attitude variations of ECA have been modeled while holding the speaking turn, but not when being the listener. The same methodology can be used to design an agent conveying attitude variations when listening to its interaction partner. Attitudes are expressed through both verbal and non-verbal behaviors. Several works showed that combining both non-verbal and verbal modalities may lead to better attitude recognition [26], [24]. Our model has only focused on the non-verbal behavior for attitude expression. In [34], the authors proposed a model to express attitudes verbally. Aspects such as the length of sentences, the variety of vocabulary or the quantity of pronouns can be taken into account in order to characterize the perceived attitude of a sentence. One can extend this type of verbal models, such as [34], and combine it with our model in order to allow ECA to express attitude variations through both verbal and non-verbal behaviors.

On another note, we did not consider the intensity of attitude variation (small, large, etc.). In the future, we intend to investigate how the perception of an attitude variation is influenced by the intensity of this variation.

To our surprise, the extracted patterns expressing "no attitude" (attitude value around zero) are evaluated as conveying friendliness. This could be caused by the annotation schema that was used to annotate the perception of attitudes. Annotators have continuously indicated the values (between -1 and 1) of the perceived attitudes. As reported in [73], a drawback with continuous annotation is the low degree of reliability between annotators. To address this issue, the same source proposed AffectRank: a rank-based annotation tool. This annotation approach should yield significantly less noise and higher inter-annotator agreement [74], [75]. We plan to use AffectRank to better annotate the perceived attitudes on the circumplex.

As noticed in our second evaluation study, the context in which is placed the agent matters. Context is to be viewed in broad sense here. For example, it can encompass the role, culture and gender of the agent. Context affects the agent's perception and needs to be taken into consideration. Furthermore, for the moment, our model varies the agent's

behavior at the signal level but not at the strategy level, that is deciding explicitly when to show an attitude variation and which one in the interaction.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant Agreements Number 769553 and Number 645378.

REFERENCES

[1] K. R. Scherer, "What are emotion? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.
[2] M. Chindamo, J. Allwood, and E. Ahlsén, "Some Suggestions for the Study of Stance in Communication," *2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pp. 617–622, 2009.
[3] D. R. Carney, J. A. Hall, and L. S. LeBeau, "Beliefs about the nonverbal expression of social power," *Journal of Nonverbal Behavior*, vol. 29, no. 2, pp. 105–123, 2005.
[4] J. K. Burgoon, D. B. Buller, J. L. Hale, and M. A. Turck, "Relational Messages Associated With Nonverbal Behaviors," *Human Communication Research*, vol. 10, no. 3, pp. 351–378, 1984.
[5] B. A. Burgoon, J. K. and Le Poire, "Nonverbal cues and interpersonal judgments : Participant and observer perception of intimacy , dominance , composure and formality," *Communication Monographs*, vol. 15, no. 3, pp. 105–124, 1999.
[6] R. Gifford and D. W. Hine, "The role of verbal behavior in the encoding and decoding of interpersonal dispositions," pp. 115–132, 1994.
[7] A. Mignault and A. Chaudhuri, "The many faces of a neutral face: Head tilt and perception of dominance and emotion," *Journal of Nonverbal Behavior*, vol. 27, no. 2, pp. 111–132, 2003.
[8] R. Gifford, "Mapping nonverbal behavior on the interpersonal circle," *Journal of Personality and Social Psychology*, vol. 61, no. 2, pp. 279–288, 1991.
[9] C. Debras and A. Cienki, "Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking," *2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, no. August, pp. 932–937, 2012.
[10] T. Stivers, "Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation," *Research on Language and Social Interaction*, vol. 41, no. 1, pp. 31–57, 2008.
[11] J. Argyle, M., Dean, "Eye contact, distance and affiliation," *Sociometry,*, no. 28, pp. 289–304, 1965.
[12] S. Duncan and D. W. Fiske, *Face-to-Face interaction: Research, methods and theory.* L. Erlbaum Associates, 1977.
[13] J. A. Hall, E. J. Coats, and L. S. LeBeau, "Nonverbal behavior and the vertical dimension of social relations: A meta-analysis," *Psychological Bulletin*, vol. 131, no. 6, pp. 898–924, 2005.
[14] J. K. Burgoon and J. L. Hale, "The fundamental topoi of relational messages," in *Communication Monographs*, 1984, pp. 193–214.
[15] D. R. Heise, *Surveying cultures: Discovering shared conceptions and sentiments.*, 23rd ed. John Wiley & Sons, 2010.
[16] T. Leary, *Interpersonal Diagnosis of Personality: Functional Theory and Methodology for Personality Evaluation*. New York: Ronald Press, 1957.
[17] Kiesler, *Contemporary Interpersonal Theory and Research, Personality, Psychopathol- ogy and Psychotherapy*, 1996.
[18] N. Bee, S. Franke, and E. Andŕe, "Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents," *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, 2009.
[19] P. Ekman and W. V. Friesen, "Felt , False , and Miserable Smiles," *Journal of Nonverbal Behavior*, vol. 6, no. 4, pp. 238–252, 1982.
[20] J. D. Mcdaniel and M. Si, "Length of Smile Apex as Indicator of Faked Expression," in *Affective Agents Workshop at IVA*, 2014, pp. 25–32.
[21] D. Keltner, "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame." *Journal of Personality and Social Psychology*, vol. 68, no. 3, pp. 441–454, 1995.
[22] J. Cassell, T. Bickmore, H. Vilhjálmsson, and H. Yan, "More Than Just a Pretty Face: Affordances of Embodiment," in *15th International Conference of Intelligent Virtual Agents (IVA 2015)*, 2000, pp. 52–59.
[23] B. Nojavanasghari, T. Baltru, C. E. Hughes, and L.-p. Morency, "The Future Belongs to the Curious : Towards Automatic Understanding and Recognition of Curiosity in Children," no. September, pp. 16–22.
[24] M. Chollet, M. Ochs, and C. Pelachaud, "A Methodology for the Automatic Extraction and Generation of Non-Verbal Signals Sequences Conveying Interpersonal Attitudes," *IEEE Transactions on Affective Computing*, no. September, pp. 1–1, 2017.
[25] B. Nojavanasghari and C. E. Hughes, "Exceptionally Social : Design of an Avatar-Mediated Interactive System for Promoting Social Skills in Children with Autism," no. May, 2017.
[26] N. Bee, C. Pollock, E. André, and M. Walker, "Bossy or Wimpy: Expressing Social Dominance by Combining Gaze and Linguistic Behaviors," in *Proceedings of Intelligent Virtual Agents: 10th International Conference, (IVA 2010)*, Philadelphia, PA, USA, pp. 265–271.
[27] C. Straßmann, A. R. von der Pütten, and R. Yaghoubzadeh, "The effect of an intelligent virtual agent' s nonverbal behavior with regard to dominance and cooperativity Dominant nonverbal behavior," *Proceedings of International Conference on Intelligent Virtual Agents*, 2016.
[28] A. Cafaro, H. H. Vilhjálmsson, T. Bickmore, D. Heylen, K. R. Jóhannsdóttir, and G. S. Valgarosson, "First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7502 LNAI, pp. 67–80, 2012.
[29] A. Cafaro, N. Glas, and C. Pelachaud, "The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions," *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*, pp. 911–920, 2016.
[30] Z. Kasap, M. B. Moussa, P. Chaudhuri, and N. Magnenat-Thalmann, "Making them remember - Emotional virtual characters with memory," *IEEE Computer Graphics and Applications*, vol. 29, no. 2, pp. 20–29, 2009.
[31] M. Ochs, N. Sabouret, and V. Corruble, "Simulation of the Dynamics of Nonplayer Characters ' Emotions and Social Relations in Games," no. January, 2010.
[32] F. Pecune, M. Ochs, S. Marsella, and C. Pelachaud, "SOCRATES: from SOCial Relation to ATtitude ExpressionS," *Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, pp. 921–930, 2016.
[33] B. Ravenet, M. Ochs, and C. Pelachaud, "From a user-created corpus of virtual agent's non-verbal behavior to a computational model of interpersonal attitudes," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8108 LNAI, pp. 263–274, 2013.
[34] Z. Callejas, B. Ravenet, M. Ochs, and C. Pelachaud, "A computational model of social attitudes for a virtual recruiter," *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, vol. 1, pp. 93–100, 2014.
[35] R. Srikant and R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance Improvements," in *5th International Conference on Extending Database Technology: Advances in Database Technology (EDBT '96)*, 1996, pp. 3–17.
[36] M. Chollet, M. Ochs, and C. Pelachaud, "Mining a multimodal corpus for non-verbal behavior sequences conveying attitudes," in *9th International Conference on Language Resources and Evaluation (LREC 2014)*, 2014, pp. 3417–3424.
[37] R. Niewiadomski, S. J. Hyniewska, and C. Pelachaud, "Constraint-based model for synthesis of multimodal sequential expressions of emotions," *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 134–146, 2011.
[38] M. S. Magnusson, "Discovering hidden time patterns in behavior: T-patterns and their detection." *Behavior research methods, instruments, & computers*, vol. 32, no. 1, pp. 93–110, 2000.
[39] S. With and S. Kaiser, "Sequential patterning of facial actions in the production and perception of emotional expressions," *Swiss Journal of Psychology*, vol. 70, no. 4, pp. 241–252, 2011.
[40] M. Guillame-Bert and J. L. Crowley, "Learning Temporal Association Rules on Symbolic Time Sequences," *Proceedings of the 4th Asian Conference on Machine Learning*, vol. 25, pp. 159–174, 2012.
[41] R. Zhao, T. Sinha, A. W. Black, and J. Cassell, "Socially-Aware Virtual Agents : Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior," in *16th International Conference of Intelligent Virtual Agents*, 2016, pp. 218–233.
[42] T. Janssoone, "Using temporal association rules for the synthesis of embodied conversational agents with a specific stance," in *16th*

International Conference of Intelligent Virtual Agents, Los Angeles, CA, USA, 2016, pp. 175–189.
[43] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schröder, "The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 5–17, 2012.
[44] D. Fricker, H. Zhang, and C. Yu, "Sequential pattern mining of multimodal data streams in dyadic interactions," in *IEEE International Conference on Development and Learning (ICDL 2011)*, 2011.
[45] C. Yu, M. Scheutz, and P. Schermerhorn, "Investigating multimodal real-time patterns of joint attention in an HRI word learning task," in *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010)*, 2010, pp. 309–316.
[46] H. Zhang, D. Fricker, T. G. Smith, and C. Yu, "Real-time adaptive behaviors in multimodal human-avatar interactions," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI 2010)*, 2010, pp. 1–8.
[47] H. Zhang and M. J. Boyles, "Visual exploration and analysis of human-robot interaction rules," in *Visualization and Data Analysis (SPIE 8654)*, 2013.
[48] M. Chollet, M. Ochs, and C. Pelachaud, "From Non-verbal Signals Sequence Mining to Bayesian Networks for Interpersonal Attitudes Expression," in *14th International Conference on Intelligent Virtual Agents (IVA 2014)*, 2014, pp. 120–133.
[49] N. Shaji and S. Izudheen, "Bio Sequence Data Mining : a Survey," *Asian Journal of Computer Science And Information Technology*, vol. 4, no. 3, pp. 21–24, 2014.
[50] B. Nicolas, "Discovering Linguistic Patterns using Sequence Mining," in *Proceedings of the 13th international conference on Computational Linguistics and Intelligent Text Processing*, 1012.
[51] R. S. Rakesh Agrawal, "Fast Algorithms for Mining Association Rules," *The Annals of pharmacotherapy*, vol. 42, no. 1, pp. 62–70, 1994.
[52] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "PrefixSpan,: mining sequential patterns efficiently by prefix-projected pattern growth," in *17th International Conference on Data Engineering (ICDE 2001)*, 2001, pp. 215–224.
[53] T. Guyet and R. Quiniou, "Mining temporal patterns with quantitative intervals," in *IEEE International Conference on Data Mining Workshops (ICDM Workshops 2008)*, 2008, pp. 218–227.
[54] G. Ruan, H. Zhang, and B. Plale, "Parallel and Quantitative Sequential Pattern Mining for Large-scale Interval-based Temporal Data," in *2014 IEEE International Conference on Big Data (BigData 2014)*, 2014, pp. 32–39.
[55] S. Dermouche and C. Pelachaud, "Sequence-Based Multimodal Behavior Modeling for Social Agents," in *roceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*, Tokyo, Japan, 2016.
[56] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: A professional framework for multimodality research," *Proceedings Language Resources and Evaluation*, pp. 1556–1559, 2006.
[57] R. Cowie, G. McKeown, and E. Douglas-Cowie, "Tracing Emotion: An Overview," *International Journal of Synthetic Emotions*, vol. 3, no. 1, pp. 1–17, 2012.
[58] S. Mariooryad and C. Busso, "Correcting time-continuous emotional labels by modeling the reaction lag of evaluators," *IEEE Transactions on Affective Computing*, vol. 6, no. 2, pp. 97–108, 2015.
[59] F. Pecune, A. Cafaro, M. Chollet, P. Philippe, and C. Pelachaud, "Suggestions for Extending SAIBA with the VIB Platform," in *Workshop on Architectures and Standards for IVAs, held at the '14th International Conference on Intelligent Virtual Agents (IVA 2014)*. Boston, MA, USA: Bielefeld eCollections, 2014, pp. 16–20.
[60] M. Argyle, *Bodily Communication*. Methuen, 1988.
[61] J. V. Bradley, "Complete counterbalancing of immediate sequential e ects in a latin square design," *American Statistical Association*, vol. 53, no. 282, pp. 525–528, 1958.
[62] K. D. Locke and E. J. Adamic, "Interpersonal circumplex vector length and interpersonal decision making," *Personality and Individual Differences*, vol. 53, no. 6, pp. 764–769, 2012.
[63] J. S. Wiggins, "A psychological taxonomy of trait-descriptive terms: The interpersonal domain." *Journal of Personality and Social Psychology*, no. 37, pp. 395–412, 1979.
[64] B. Ravenet, A. Cafaro, B. Biancardi, M. Ochs, and C. Pelachaud, "Conversational behavior reflecting interpersonal attitudes in small group interactions," in *15th International Conference of Intelligent Virtual Agents (IVA 2015)*, 2015, pp. 375–388.
[65] R. Op Den Akker, M. Bruijnes, R. Peters, and T. Krikke, "Interpersonal stance in police interviews: Content analysis," *Computational Linguistics in the Netherlands Journal*, vol. 3, pp. 193–216, 2013.
[66] J. S. Wiggins, P. Trapnell, and N. Phillips, "Psychometric and Geometric Characteristics of the Revised Interpersonal Adjective Scales (IAS-R)," pp. 517–530, 1988.
[67] V. Y. Yzerbyt, N. Kervyn, and C. M. Judd, "Compensation versus halo: The unique relations between the fundamental dimensions of social judgment," *Personality and Social Psychology Bulletin*, vol. 34, no. 8, pp. 1110–1123, 2008.
[68] K. D. Locke, "Circumplex Measures of Interpersonal Constructs," *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*, no. March, pp. 313–324, 2012.
[69] B. Knutson, "Facial expressions of emotion influence interpersonal trait inferences," *Journal of Nonverbal Behavior*, vol. 20, no. 3, pp. 165–182, 1996. [Online]. Available: http://link.springer.com/10.1007/BF02281954
[70] L. Tiedens, P. Ellsworth, and B. Mesquita, "Stereotypes about sentiments and status: Emotional expectations for high- and low-status group members," *Personality and Social Psychology Bulletin*, vol. 26, no. 5, pp. 560–574, 2000.
[71] H. Vilhjalmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. Van Welbergen, and R. J. Van Der Werf, "The behavior markup language: Recent developments and challenges," *Lecture Notes in Artificial Intelligence*, vol. 4722, no. 1, pp. 99–111, 2007.
[72] D. Heylen, S. Kopp, S. C. Marsella, C. Pelachaud, and H. Vilhjálmsson, "The next step towards a function markup language," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5208 LNAI, pp. 270–280, 2008.
[73] G. N. Yannakakis, "The Ordinal Nature of Psychophysiology," in *Proceedings of the 5th International Conference on Physiological Computing Systems*, Seville, Spain, 2018, pp. 248–255.
[74] G. N. Yannakakis and H. P. Martinez, "Grounding truth via ordinal annotation," *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, pp. 574–580, 2015.
[75] G. N. Yannakakis, R. Cowie, and C. Busso, "The Ordinal Nature of Emotions," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2017, pp. 248–255.

Soumia Dermouche is researcher at Kiliba company in France. Her research activities are focused on leveraging the dynamics of non-verbal behaviors for modeling social phenomena such as attitude and engagement in human-agent interaction.

Catherine Pelachaud is a research director at ISIR-CNRS laboratory at Sorbonne Université. Her research interests include embodied conversational agent, non-verbal communication (face, gaze, and gesture), expressive behaviors, and socio-emotional agents.