

User Attitude Towards an Embodied Conversational Agent: Effects of the Interaction Mode

Nicole Novielli, Fiorella de Rosis and Irene Mazzotta

Department of Informatics

University of Bari, Italy

Via Orabona 4, 70126 Bari, Italy

{novielli, derosis, mazzotta}@di.uniba.it

Abstract

We describe how the interaction mode with an Embodied Conversational Agent (ECA) affects the users' perception of the agent and their behavior during interaction, and propose a method to recognize the social attitude of users towards the agent from their verbal behavior. A corpus of human-ECA dialogues was collected with a Wizard of Oz study in which the input mode of the user moves was varied (written versus speech-based). After labeling the corpus, we evaluated the relationship between input mode and social attitude of users towards the agent. The results show that, by increasing naturalness of interaction, spoken input produces a warmer attitude of users and a richer language; this effect is more evident for users with a background in humanities. Recognition of signs of social attitude is needed for adapting the ECA's verbal and nonverbal behavior.

keywords: user-centered design; natural language user interfaces; evaluation of artificial agents; health care applications.

Biographical notes

Nicole Novielli is a PhD student in Computer Science in the Department of Informatics, University of Bari. Her interests are in the area of recognition of affective and cognitive states from language.

Fiorella de Rosis is a Full Professor of Computer Science in the Department of Informatics, University of Bari. She teaches Intelligent Interfaces at the Master-level degree in Informatics. She is a Member of ACM, ISRE (the International Society for Research on Emotions) and HUMAINE. Her research interests are in the area of intelligent interfaces, user modeling and adaptation and affective computing.

Irene Mazzotta is a Research Assistant in the Department of Informatics, University of Bari. Her interests are in the area of dialogue simulation and user-adapted persuasion.

Corresponding author:

Nicole Novielli

Dipartimento di Informatica

Università degli Studi di Bari

Via Orabona 4

70126 Bari, Italy

novielli@di.uniba.it

1. Introduction

In the majority of existing applications, users are enabled to interact with Embodied Conversational Agents (ECAs) with keyboard and mouse. However, while using a keyboard to communicate with an agent that talks and simulates human-like expressions is quite unnatural, a speech-based user input can be seen as a more natural way to interact with a human-like interlocutor. In addition, spoken interaction is likely to become more common in the near future. Humans proved to align themselves in conversations by matching their nonverbal behavior and word use: they instinctively converge in the number of words used by turn and in the selection of terms belonging to ‘social/affect’ or ‘cognitive’ categories (Niederhoffer and Pennebaker, 2002). ECAs should be able to emulate this ability; understanding whether and how the interaction mode influences the user behavior and defining a method to recognize relevant aspects of this behavior is therefore important in establishing how the ECA should adapt to the situation.

Among the various media that are involved in human-ECA interaction (gaze, face, gestures, body posture etc), we focused our study on language. Our long-term goal is to create an agent that informs, persuades and engages a human interlocutor in a conversation about healthy dieting. We expect that the attitude towards the ECA will not be the same for all users and that it will vary mainly according to their goal (Walton, 2006). We also hypothesize that the way this attitude is displayed will vary during the dialogue, depending on the topic discussed in every phase. For this reason, we aim at recognizing the particular ways in which users display their interpersonal stance towards the ECA, by looking at what we call ‘signs’ of this attitude so as to adapt the agent’s behavior accordingly (Mazzotta et al., 2007). By using this information, the overall dialogue strategy and the agent’s plan will be adapted, to meet users’ goals and expectation. By looking at the linguistic signs of the users’ attitude, short term behavior will be tailored as well. Finally, the language style, word use and facial expressions of the ECA will be coordinated with the overall attitude of the users. This paper builds on previous research in two areas: on one hand, studies which investigate the factors influencing the users’ behavior during the interaction with animated characters; on the other hand, studies aimed at defining methods to recognize affective states from language features. As we anticipated, among the various features that characterize the user behavior, we considered a particular aspect of interpersonal stance that we named *social attitude*. In previous work, we proposed a method based on keyword-spotting techniques to recognize this attitude in text-based interaction (de Rosis et al., 2006). This paper describes an extension of that work, by studying how the interaction mode with the ECA (via keyboard and mouse vs. via microphone and touch screen) influences the user behavior, and whether the accuracy in recognizing this attitude may be increased by means of statistical language processing methods (Charniak, 1993). In Section 2, we clarify what we mean by ‘social attitude’ and position our work in the vein of ongoing related research. Section 3 describes the Wizard of Oz experimental study with which a corpus of dialogues was collected and how these data were analyzed. In Section 4, we propose a method to recognize signs of social attitude in individual dialogue moves and apply it to the corpus. Some final remarks discuss the results obtained and the problems still open (Section 5).

2. Related work

Affective states vary in their degree of stability, ranging from long-standing features (personality traits) to more transient ones (emotions). ‘Interpersonal stance’ is in the middle of this scale (Scherer et al., 2004): it is initially influenced by individual features like personality, social role and relationship between the interacting people but may be changed, in valence and intensity, by episodes occurring during interaction. After the concept of ‘socially intelligent agents’ was first introduced (Dautenhan, 1998), some variants of interpersonal stance of humans towards technology were studied under different names, in various research projects. Some researchers talk about ‘engagement’ to denote “*the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake.*” (Sidner and Lee, 2003) or “*how much a participant is interested in and attentive to a conversation*” (Yu et al., 2004). The term ‘social presence’ (Rettie, 2003) was employed to denote “*the extent to which the communicator is perceived as ‘real’*” (Polhemus et al., 2001) or, more specifically, “*the extent to which individuals treat embodied agents as if they were other real human beings*” (Blascovich, 2002). Bailenson et al. (2005b) clarified the difference between these two definitions by distinguishing *perception* of ECAs from *social response* to them. In this paper we will focus on the user social response to the ECA by distinguishing between ‘warm’ and ‘cold’ *social attitude*. In particular, we will refer to the concept of interpersonal *warmth* that was introduced by Andersen and Guerrero (1998) to denote “*the pleasant, contented, intimate feeling that occurs during positive interactions with friends, family, colleagues and romantic partners.*”

A large variety of nonverbal markers of interpersonal stance have been proposed: body distance, memory, likeability, physiological data, task performance, self-report and others (Bailenson et al., 2005a; Bailenson et al., 2005b). Forms of expression with language in human-human communication (Andersen and Guerrero, 1998) or online discussions (Polhemus, 2001; Swan, 2002) have been reported as well. However, experimental studies about how interaction modality with embodied agents influences the verbal behavior of users are still quite limited. In an experiment comparing spoken with written data input on a simple task, Oviatt et al. (1994) found longer utterances, larger variability of lexical content, less frequent use of abbreviations and signs and larger syntactic ambiguity in spoken than in written data. The same research group analyzed how primary school children adapt their language when interacting with animated characters (Oviatt and Adams, 2000); they found that several forms of idiosyncratic linguistic constructions were employed: invented words, incorrect lexical selections, ill-formed grammatical constructions, mispronounced or exaggerated articulations and questions about animated characters’ personal characteristics. Similar verbal behaviors were found by Zhang et al. (2006) in developing a language-based affect detection module to control an automated virtual actor. In her studies on online teaching, Swan (2002) demonstrated that “*media with few affective communication channels (such as text-based computer-mediated communication) have less social presence than media with a greater number of affective communication channels.*”

For a long time, recognition of affective states mainly focused on *basic emotions* (Ekman, 1999) such as joy, fear, anger etc. Only more recently we saw the birth of

studies aimed at recognizing the kinds of affective states that are more likely to occur in human-machine interaction such as frustration, boredom, confusion, effort and others¹.

Rather than focusing on recognition of emotions, research in this paper is an attempt to consider the users' *social attitude* towards an artificial agent: starting from a corpus of Wizard of Oz dialogues we study how this attitude is influenced by the interaction mode (text vs speech) and how it may be recognized with statistical language processing methods.

(1) Our study

Our study is focused on the effects (on the dialogue) of the interaction mode (text-based vs. speech-based) rather than of the ECA's attributes (as, for instance, in Nass et al., 2000; Bickmore and Cassell, 2005). The ECA's competence and the modality with which its messages were rendered stayed unvaried during all the experiments while we varied the communication channel enabled to the subjects. We considered, in particular, the following questions:

(Q1): Is the subjects' *social attitude* towards our ECA influenced by the interaction mode? That is, does the input mode (written input via keyboard vs spoken input via microphone) affect the social attitude displayed by subjects?

(Q2): Can the various forms in which social attitude is expressed in dialogues be recognized through *language analysis methods* (de Rosis et al., 2007)?

3.1. Study design

Research proved that users tend to comply more with artificial agents whose appearance matches the sociability required in the jobs they simulate (Goetz et al., 2003) and, in particular, that female agents are preferred when acting as therapists or medical advisors (Zimmerman et al., 2005): we therefore employed in our experiment a 3D realistic agent shaped as a youthful woman named *Valentina*. The agent was implemented by integrating the Haptik player² with an Italian text-to-speech system (TTS) by Loquendo³, and used facial expressions and head movements to enrich its speech acts with affective and mental state meanings.

A corpus of dialogues with this ECA was collected in a between-subject study with a Wizard of Oz tool⁴ (Clarizio et al., 2006). Sixty graduated students (age 21-28) were

¹ For a description of some very recent studies in this domain, see the International Journal of 'User Modeling and User-Adapted Interaction', Special Issue on "Affective User Modeling", in press.

² <http://www.haptik.com>

³ <http://www.loquendo.com>

⁴ Wizard of Oz studies are a popular evaluation method in which subjects interact with what they believe to be an implemented system while a 'wizard' interprets their moves and selects and forwards the system answer from a remote server (Dahlback et al., 1993). When used to simulate dialogue systems, they enable observing the linguistic

involved in the study, thirty for every interaction mode: in the written-input setting, users could interact with the ECA with keyboard and mouse; in the spoken-input condition, the ECA was displayed on a touch screen and users used a microphone to talk to it and a touch-screen to send other commands (as shown in Figure 1). The two groups of subjects were balanced for gender and university curricula (computer science or humanities). To insure uniformity of experimental conditions throughout the whole study, the wizard (the same person for every dialogue) was trained so as to apply a dialogue plan specified in a document. In particular, at every dialogue step, the wizard selected its next move from among those available on his server-side, by following the established dialogue plan: after initial self-introduction, information about the eating habits of the subject was collected, before providing tailored information and suggestions about healthy eating and justifying them. As a result, the ECA had exactly the same visual and audio behavior in the two modes, and also the set of dialogue moves available to the wizard was the same (78 moves overall). These moves were organized into categories: self-introduction (in which the agent introduced itself and described its role), questions about the subjects' eating habits (what they used to eat, what they liked, etc), suggestions (advantages or disadvantages of various meal components or combinations), general comments ('*Good question!*', '*You are right!*') and farewell. To ensure the believability of the Wizard of Oz experiment, the tool architecture was designed so as to allow the wizard to successfully simulate the real-time reaction of a working system: (e.g. the wizard could read/listen the subject move while he was still typing/formulating it; in the wizard interface, the set of moves were organized and presented according to the mentioned categories, so as to be instantaneously retrieved by the wizard; etc.)

FIGURE 1 ABOUT HERE

Subjects' moves were completely unconstrained: they could just answer the agent's questions and hear its suggestions or could take the initiative to make questions and comments of any kind. Overall, 1614 dialogue moves were collected: 712 in the dialogues *with written input* (D1 corpus) and 902 in those with *spoken input* (D2 corpus). At the end of interaction, subjects were asked to fill out an online questionnaire (Figure 2) in which they could express, on a five-point Likert scale, their evaluation of the information received during interaction and of the ECA, as well as whether they would have preferred to receive information about healthy eating from a conventional information system rather than an ECA.

FIGURE 2 ABOUT HERE

To build a corpus of data to analyze, dialogues with written input were automatically stored during the experiment while those with spoken input were manually transcribed from audio files.

behavior of users and collecting a corpus which can be very useful to set up and test language analysis methods.

3.2 Corpus annotation

We defined a mark-up language for the user moves after carefully examining our corpus and considering suggestions from the studies about verbal expression of social attitude that we cited in Section 2: Swan (2002) proposes a coding schema for analyzing social communication in text-based interaction which employs *affective, cohesive and interactive indicators*. Similar indicators have been suggested by Polhemus et al. (2001):

- personal address and acknowledgement (using the name of the persons to which one is responding, restating their name, agreeing or disagreeing with them),
- feeling (using descriptive words about how one feels),
- paralanguage (features of language which are used outside of formal grammar and syntax, which provide additional enhanced, redundant or new meanings to the message),
- humor,
- social sharing (sharing of information not related to the discussion),
- social motivators (offering praise, reinforcement and encouragement),
- value (set of personal beliefs, attitudes),
- negative responses (disagreement with another comment),
- self-disclosure (sharing personal information).

Finally, other indicators have been proposed by Andersen and Guerrero (1998), whose definition of interpersonal warmth we refer to when talking about warm social attitude:

- sense of intimacy (use of a common jargon),
- attempt to establish a common ground,
- humor,
- benevolent or polemic attitude towards the system failure,
- interest to protract or close the interaction.

In defining the signs to include in our language, we also considered the form of adaptation we wanted to implement in the agent's behavior (Carofiglio et al., 2005). Dynamic recognition of individual signs is not simply aimed at evaluating the overall polarity of social attitude shown by the user: evidence about every linguistic sign (see Table 1) observed during the dialogue enables adapting the short-term agent's plan accordingly: for example, if the user tends to talk about herself, in its following moves the ECA will use this information to provide more appropriate suggestions. The overall social attitude of the user will be inferred dynamically from the history of signs recognized during the dialogue (Liu and Maes, 2004) at the move level to adapt the ECA's language style, voice and facial expression.

Table 1 describes the signs of social attitude included in our markup language, with their definitions and some examples of dialogue exchanges.

Table 1 ABOUT HERE

Three independent raters were asked to annotate the corpora D1 and D2: these were presented separately, with dialogue exchanges in random order. Inter-rater agreement was measured for all the signs excluding humor (of which we found few, although quite interesting, cases in our dialogues); the results of ‘majority agreement’ rates (when 2 out of 3 raters agreed) are illustrated in Table 2.

This table shows that the distribution of signs was quite unequal, in both interaction modes: talks about self, questions about the agent and colloquial style were the most frequent of them. Some authors criticize the percent agreement estimates of interrater reliability, on the grounds that they do not account for chance agreement among coders; instead, they prefer Cohen’s kappa, which is a chance-corrected measure (Di Eugenio and Glass, 2004): we report both measures in the table, which shows that overall, our markup language proved to be robust (Di Eugenio, 2000).

Table 2 ABOUT HERE

3.3. Relationship between interaction mode and social attitude.

In this section we provide some results to partially address question Q1: *Is the subjects’ social attitude towards our ECA influenced by the interaction mode?* More results about the effect of interaction mode on users’ linguistic behavior will be provided in section 4. Here we provide results about the analysis of some quantitative features of the dialogues, which show that the input mode influenced significantly the dialogue characteristics: the average number of moves per dialogue was higher in spoken input (30.7 vs. 23.6, two-sided $p=.03$), as well as the average number of characters per move (81.1 vs. 47.7, $p=.005$); move length was also influenced by the subjects’ background (86.2 in humanities vs. 43.4 in computer science, $p=.001$).

Table 3 ABOUT HERE

We computed, for every subject, the *proportion of moves in the dialogue which, according to a majority agreement among raters, displayed at least one sign of social attitude*. This proportion can be considered as an overall index of social attitude: the more frequently users display signs of social attitude in their moves, the more likely they can be considered as displaying a social attitude towards the ECA. The average value of this variable was .45: this means that almost half of the moves in our corpus (D1 + D2) included at least one sign of social attitude. Multiple regression analysis (Table 3) shows the factors which influence this index: in order of importance, background in humanities, dialogue length and move length. As longer dialogues occur in the spoken-input mode, this means that a higher percentage of moves that were classified as ‘social’ by our raters occurred in this mode.

These results enable us to answer the first question of our study.

Answer to Q1: the subjects' social attitude towards the ECA is influenced by the interaction mode: spoken input entails longer dialogues, both in number of moves and in move length, with a larger percentage of social moves.

This finding agrees with results of the study by Oviatt et al. (1994), according to which spoken input entails longer utterances than written input: as we will see later in this paper, this similarity between our studies extends to variations in the language employed in the two modalities. The finding does not explain, however, whether subjects using the spoken input mode displayed a warmer social attitude only because, with the increased length of the move, the probability of introducing some sign of this attitude increased as well, or because they really behaved more socially. Moreover, according to the final questionnaire data, the average evaluation of the information received and the ECA was 3.06; this means that the subjects felt, on average, that the information received was worthwhile and that the ECA was an acceptable means to provide it: *no* subject declared that they would have preferred interacting with a conventional information system rather than an ECA, in that domain.

We then tested whether there was any relationship between questionnaire evaluation and percentage of social moves, that is between what Bailenson et al. (2005a) called 'perception of the ECA' and social attitude. A simple regression model shows, quite surprisingly, that no such relationship seems to exist (R-square = .0049, st. error of estimate = .29). This finding agrees with results of other studies according to which questionnaires would not be as sensitive as behavioral measures in quantifying differences in affective responses to virtual agents (Bailenson et al., 2005a; Bailenson et al., 2005b; Picard and Daily, 2005; Höök et al., 2005). It seems to provide a proof in favor of the thesis that users' behavior when interacting with ECAs (the time they spend in interaction, their propensity to interact with them etc.) is not associated with their evaluation of the agent, but rather with their individual dispositions, which should be considered when adapting the conversational strategy.

4. Recognizing signs of social attitude in individual dialogue moves

The second part of our research was aimed at answering question *Q2*, whether social attitude can be recognized with language analysis methods. In the majority of studies on affect recognition, language analysis is seen as complementary to prosodic analysis rather than as a method per se (Lee et al., 2002; Ang et al., 2002; Litman et al., 2003; Batliner et al., 2003; Devillers and Vidrascu, 2006). Analysis of affect valence or opinion polarity in texts is the domain of 'sentiment analysis' (Wilson et al., 2005) and 'points of view' recognition (Liu and Maes, 2004), while textual emotion estimation is aimed at recognizing specific emotional states (Neviarouskaya et al., 2007).

Methods applied in the recognition of affective features range from simple keyword spotting to more sophisticated approaches. Machine learning was applied to automatically infer, from text samples, what indicators are useful to categorize them (Pang et al., 2002; Cunningham et al., 1997). In other cases, linguistic or semantic knowledge was used to classify words into *ad hoc* categories and to compute an overall score for the text subsequently, based on the 'bag-of-words' occurrence or frequency (Whitelaw et al., 2005). In statistical language modeling, the probability of a sentence is

decomposed into a product of conditional probabilities of the words in the sentence given their ‘history’ (Rosenfeld, 2000). In Latent Semantic Analysis (LSA) by Landauer and Dumais (1997), an index of term-document similarity is computed from a term-by-document matrix; to reduce the ‘sparse data’ problem, word sequences can be classified into linguistic categories, and category-by-document matrices are then analyzed rather than term-by-document ones. The method we applied shares with LSA the idea of working on categories-by-document matrices but introduces sign-specific categories.

1. Method

Individual user moves are analyzed to recognize the presence of *all* signs of social attitude described in Table 1 in probabilistic terms. A sign can be displayed in a move through some ‘linguistic cues’: these may range from single words to short word sequences or sentence patterns. For example, as shown in the examples in Table 4, a ‘Friendly self-introduction’ may include word sequences like ‘nice to meet you’, ‘my name is’ or ‘ciao’; ‘Talks about self’ may include first person pronouns or verbs like ‘I’ or ‘I eat’, etc. We classified the set of linguistic cues that are typical of the seven signs we considered in our analysis into 32 ‘linguistic categories’, according to syntactic or semantic criteria. Table 4 shows these linguistic categories (column 2), some examples of their content (column 3) and their relationship with the signs of social attitude (column 1). For example: a ‘nice to meet you’ belongs to the category of ‘Greeting’, which is associated with the sign ‘Friendly self-introduction’. The linguistic cues associated to every sign of social attitude have been defined according to both the theories described in section 2.

This table shows that linguistic categories are not necessarily ‘salient’ (Lee et al., 2002) for only a single sign; also, they are not necessarily disjoint. For example, the ‘Ciao’ category is associated with both a ‘Friendly self-introduction’ and a ‘Friendly farewell’. Therefore, a sign can be recognized in the text of a user move only in conditions of uncertainty. Sign-category associations are employed by our linguistic analyzer to check, for every sign s_j , whether any of the word sequences in the categories (c_h, c_k, \dots, c_z) that are associated with s_j is present in the text. The result of this analysis is a binary vector $V(c_h, c_k, \dots, c_z)$ that describes whether the move includes any item in the categories that are relevant for the sign s_j . For example, when applied to recognize possible signs of Friendly Farewell in the move: “*Bye, I would like to thank you for spending your virtual time with me*”, the linguistic analyzer examines the categories ‘Expression of farewell’, ‘Thanking’ and ‘Ciao’, and outputs the vector (1,1,0).

 Table 4 ABOUT HERE

A Bayesian classifier (de Rosis et al., 2007) then computes the probability of observing the sign of social attitude s_j in the considered move, given the output of the linguistic analyzer $V(c_h, c_k, \dots, c_z)$. This probability value is computed from $P(s_j)$, the prior probability that s_j occurs in the text, $P(V(c_h, c_k, \dots, c_z))$, the prior probability of the combination of values in the vector, and $P(V(c_h, c_k, \dots, c_z) | s_j)$, the probability that this combination of values occurs in a move displaying s_j :

$$(1) P(s_j | V(c_h, c_k, \dots, c_z)) = \frac{P(V(c_h, c_k, \dots, c_z) | s_j) * P(s_j)}{P(V(c_h, c_k, \dots, c_z))}$$

In the previous example:

$$(2) P(ffwell | (1,1,0)) = \frac{P((1,1,0) | ffwell) * P(ffwell)}{P((1,1,0))}$$

The parameters of the Bayesian classifier (prior probabilities) were learnt from the two corpora of dialogues. The result of Bayesian classification of a move is a vector of probabilities for every sign. A threshold is settled for these probabilities, to decide whether each sign is present in the move. Thresholds are settled after an analysis of ROC curves (Zweig and Campbell, 1993) for the various signs, which enable us to select an optimal balance between true and false positives.

To compare the language used in the two interaction modes, we proceeded as follows:

- (i) we learnt the parameters of the Bayesian classifier from the corpus D1 (written-input dialogues, 712 moves overall);
- (ii) we learnt the parameters of the Bayesian classifier from the corpus D2 (spoken-input dialogues, 902 moves overall)
- (iii) we compared the results of the two learning steps to reflect on the reasons of variations in language usage.

1. Training the classifier with the written-input corpus

Figure 3 shows the ROC curves for the various signs: false positive (FP) and true positive (TP) rates are displayed on the x and y axis respectively.

 FIGURE 3 ABOUT HERE

Table 5 shows the sensitivity and specificity of Bayesian classification when the cutoff points suggested by the ROC curves are applied.

 Table 5 ABOUT HERE

By changing the cutoff values, we trade-off sensitivity and specificity. This choice is not a technical one, but depends on the consequences of identifying a sign which was not actually expressed in a move, or missing a sign that was expressed indeed. In the *maximize sensitivity* strategy, the ECA will risk to respond with a ‘warm’ social attitude to a ‘neutral’ or ‘cold’ user behavior; in the *maximize specificity* strategy, the inverse will occur. We applied the cutoff points suggested by ROC analysis to build a ‘reasonably social’ ECA (not too cold but not too warm either). By comparing these results with those obtained in our previous study (de Rosis et al., 2006), we noticed that we had got a considerable improvement in the recognition accuracy by upgrading our simple keyword-based method: the average sensitivity increased from .63 to .90, with an average specificity staying virtually unchanged (from .93 to .90).

More than one sign may be recognized in an input move. This may be due to sentences really displaying several signs of social attitude like: “*Hi Valentina, nice to meet you! I’m curious to hear what you will suggest me!*” (‘Ciao’ ‘Friendly-self-introduction’ + ‘First-person auxiliary verb’ ‘Talks about self’) but also to misrecognition problems. The confusion matrix in Table 6 shows (in italics) the main misrecognition problems in our analysis.

 Table 6 ABOUT HERE

We identified two causes of these problems:

- *partial overlapping between word sequences in some of the linguistic categories.* For example, the sentence ‘*You are rude*’ is included in the ‘Evaluation of agent’s politeness’ category; however, the first fragment (‘*You are*’) belongs to the ‘Second-person auxiliary verbs’. The sentence is therefore classified as a ‘Negative comment’ but also as a ‘Question to the agent’. A similar problem is the cause of confounding between ‘Friendly farewell’ and ‘Question about the agent’ for the move ‘*Bye, I would like to thank you for spending your virtual time with me!*’, which includes a ‘Ciao’ and a ‘Second-person pronoun’.
- *partial overlapping between the contents of some semantic categories:* e.g., as we said, the category ‘Ciao’ is relevant to both ‘Friendly Self Introduction’ and ‘Friendly Farewell’. Therefore, ‘*Ciao, my name is Carlo*’ is recognized, at the same time, as a ‘Friendly self introduction’, a ‘Friendly farewell’ and (due to the previously mentioned problem) a ‘Talks about self’.

4.3. Comparison with the speech-input corpus

To assess the differences in the language employed in the spoken vs. written input, we first applied the Bayesian classifier algorithm learnt on corpus D1 to corpus D2. We found that sensitivity in recognizing some of the signs (friendly self-introduction, friendly farewell and talks about self) did not change, while it considerably decreased for colloquial style, and for positive and negative comments.

In looking for an explanation for this decrease of recognition power, we found that the semantic categories we had defined for every sign were still valid; however:

- (i) the contents of these categories (lists of words, word sequences or sentence patterns) had to be extended to describe the richness of language employed in speech-based interaction. There were, in this mode, a wider use of diminutives and dialect, new forms of evaluation of the agent’s competence, a much wider use of paralanguage and more ‘familiar’ expressions of positive comments;
- (ii) some of the parameters (prior and conditional probabilities) *varied*: in particular, there was a rise of probability for vectors $V(c_h, c_k, \dots, c_z)$ with few zero values. This was due to the increased length of subject moves and the wider richness of the language employed;

- (iii) some new (either positive or negative) signs of social attitude appeared, that were not displayed in the written-input form: signs of *politeness* (“Don’t mind”, “That’s kind of you!”,...), of *encouragement* (“And now, what are we going to talk about?”) but also of sarcastic paralinguistic expressions (“E mbeh?” “Eh, vabbè!”). Most of these forms may be interpreted as positive or negative comments only by considering the context in which they were uttered (previous agent move) and their acoustic properties.

According to these findings, we trained the Bayesian classifier again for the spoken corpus and we traded-off sensitivity and specificity again. Table 7 shows that sensitivity and specificity rates were a bit lower than in the written-input corpus.

Table 7 ABOUT HERE

Comparison of the confusion matrices in tables 6 and 8 shows that, when we tried to obtain sensitivity values similar to those we had in the written-input corpus (tables 5 and 7), we observed a significant decrease of specificity for some of the signs: in particular, Talks about self, Question to the agent, *Positive* and *Negative comments*. The richness of the spoken language seems to induce, as well, new causes of confounding: misrecognition increased for Negative comments, which were more frequently classified as Colloquial style (the rate increased to .44), due to the presence of some *new* sarcastic paralinguistic expressions that subjects used for expressing either real agreement or ironic negative comments. Confusion with Negative Comments was higher as well, especially for Talks about self (.47). The lower specificity for Questions about the agent produced new confounding with Talks about self (.37) and Positive comments (.22).

Table 8 ABOUT HERE

We interpreted these results as a confirmation of the other authors’ findings cited earlier, concerning the higher richness of speech-based interaction. This answers question *Q1* that we left partially open in Section 3.3, by suggesting that *the warmer social attitude displayed by users in speech-based interaction is not only due to the easier input mode (which entails an increased move length) but also to the more natural form of communication promoted by this mode*. And, of course, the richer the language, the more difficult recognition of signs of social attitude becomes. Hence our

Answer to Q2: the various forms in which users may express their social attitude towards our ECA can be recognized with Bayesian classification methods, with different degrees of accuracy, the average accuracy being higher in the case of written than in the case of spoken input.

However, recognition accuracy in spoken input can be increased by integrating linguistic analysis with analysis of prosodic features, as demonstrated in de Rosis et al. (2007).

5. Final remarks

Catrambone et al. (2004) raised the question: *“If you could ask for assistance from a smart, spoken natural language help system, would that be an improvement over an on-line reference manual?... Does it matter that the user consultant has a face and that the face can have expressions and convey a personality?”*. The authors answered positively to this question, in line with other studies and in contrast to some more negative or doubtful positions (Schneiderman and Maes, 1997). With our study, we wanted to go deeper into this question by asking: *“Would a more natural interaction mode with such a system promote a warmer attitude in users, if compared with a regular WIMP⁵ modality?”* and *“Would different users behave differently in interacting with such a kind of help systems and, if so, might these differences be recognized?”*. To our knowledge, this is the first study that considered those questions and provided an answer to them. Our evaluation of the users’ behavior was made in terms of social attitude towards the agent displayed in language. According to the results we obtained, spoken input seems to increase considerably the naturalness of access to an advice-giving ECA, by producing a warmer attitude and a higher richness of the language employed. This effect is more evident for users with a background in humanities: computer scientists tended to be more cold and formal towards the agent and to take a ‘challenging’ attitude towards it, by investigating the limits of its ‘intelligence’ with several and sometimes tricky questions (e.g.: *“How can you give suggestions about healthy dieting, if you can’t eat?”*). On the contrary, subjects with a background in humanities had a more natural attitude and made longer dialogues with more signs of social attitude.

These findings have useful implications for the design of ECAs and how they should adapt to the user characteristics. Once again, they support the idea that conversational agents that act similarly for all users is unlikely to be successful; on the contrary, ECAs should be able to recognize user attitude in order to adapt dynamically their behavior during the dialogue. Our recognition method was quite effective for some signs (‘friendly self-introduction’, ‘friendly farewell’ and ‘positive comments’), a bit less for others (‘negative comments’ and ‘talks about self’). In particular, we attempted to recognize humor, which requires much more complex language analysis methods.

While speech-based interaction warms up the attitude of users towards the agent, it does not seem to improve their evaluation of the system: this seems to depend mainly on the agent’s ability to answer appropriately to the users’ requests for information and on the ECA’s persuasion strength.

We must acknowledge several limits in our study:

- first of all, we did not assess the personality of subjects (e.g. with Myers & Briggs questionnaire⁶), in order to avoid the risk of reducing their level of cooperation. On the contrary, personality traits of users (in particular, extraversion vs. introversion) seem to affect human-ECA interaction: Bickmore and Cassell (2005) found that extroverted users preferred an ECA making some form of social dialogue more than introverted ones. This hypothesis is supported by the similarity between our signs of social attitude and some of the language features which characterize extraversion (Gill and Oberlander, 2002): according

⁵ WIMP stands for “window, icon, menu, pointing device”.

⁶ <http://www.humanmetrics.com/cgi-win/JTypes2.asp>

to that study, extraverts tend to use a ‘relaxed’ and ‘informal’ style, make less use of the first person singular pronouns and express a ‘positive affect’ more frequently; these linguistic markers are similar to those we included among our signs of social attitude;

- due to the complexity and length of conducting this kind of experiment, the corpus of dialogues was not extensive, leading to the phenomenon of ‘sparse data’, which is common in such studies;
- our subjects were similar (in age and background) to potential users of the system, and therefore the kind of social attitude they displayed towards the ECA was probably similar to the relationship a subject in need of help would establish with it; however, the study involved people who did not spontaneously ask for information about healthy eating. Therefore, we cannot be sure that our findings are completely representative of the population of users;
- the length of user-ECA interactions was quite short (from 20 to 40 minutes), while an effective support should rely on repeated encounters with subjects and would probably produce a different user-ECA relationship: probably closer, but also with a higher risk of repetitiveness and boredom;

In spite of these limits, we found several results which agree with the psycholinguistic theories of which we are aware. In the immediate future, we plan to improve the language analysis method and to strengthen its speaker-independence by linking our linguistic categories to the synsets of WordNet (Fellbaum, 1998).

Acknowledgements

This paper is dedicated to Marino, one of the most social ‘subjects’ who participated in our Wizard-of-Oz study, who died prematurely in a car accident.

The work was financed, in part, by HUMAINE, the European Human-Machine Interaction Network on Emotion (EC Contract 507422). We acknowledge Giuseppe Clarizio for developing the Wizard of Oz tool employed in our experiments and for cooperating in the data collection and Vadim Bulitko for kindly helping us in correcting our English.

References

Andersen, P.A. and Guerrero, L.K., 1998. Handbook of Communication and Emotions. Research, theory, applications and contexts, Academic Press, New York

Ang, J., Dhillon, R., Krupsky, A., Shriberg, E. and Stolcke, A., 2002. Prosody-based automatic detection of annoyance and frustration in human-computer dialog, in: J. H. L. Hansen and B. Pellom, editors, Proceedings of ICSLP, vol. 3, 2037.

Bailenson, J.N., Swinth, K.R., Hoyt, C.L., Persky, S., Dimov, A. and Blascovich, J., 2005. The independent and interactive effects of embodied agents appearance and behavior on self-report, cognitive and behavioral markers of copresence in Immersive Virtual Environments. Presence, 14, 4, 379-393.

- Bailenson, J.N., Aharoni, E., Beall, A.C., Guadagno, R.E., Dimov, A. and Blascovich, J., 2005. Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments, in: Proceedings of 7th Annual International Workshop on Presence.
- Batliner, A., Fisher, K., Huber, R., Spilker, J. and Nöth, E., 2003. How to find trouble in communication, in: *Speech Communication*, 40, pp.117-143.
- Bickmore, T., Cassell, J., 2005. Social Dialogue with Embodied Conversational Agents, in: J. van Kuppevelt, L. Dybkjaer, & N. Bernsen (Eds.), *Advances in Natural, Multimodal Dialogue Systems*. New York: Kluwer Academic.
- Blascovich, J., 2002. Social influences within immersive virtual environments, in: R. Schroeder (Eds.), *The social life of avatars*. Springer-Verlag, London, pp. 127-145.
- Carofiglio, V., de Rosis, F and Novielli, N., 2005. Dynamic User Modeling in Health Promotion Dialogues. In Proceedings of the First Conference on Affective Computing and Intelligent Interfaces, 2005.
- Catrambone, R., Stasko, J. and Xiao, J., 2004. ECA as userinterface paradigm, in: C.Pelachaud and Z Ruttkay (Eds), *From browse to trust: evaluating embodied conversational agents*, Kluwer Academic Publishers, pp. 239-267.
- Charniak, E., 1993: *Statistical Language Learning*, The MIT Press .
- Clarizio, G., Mazzotta, I., Novielli, N. and De Rosis, F., 2006. Social attitude towards a conversational character, in Proceedings of RO-MAN, pp. 2-7.
- Cunningham, S.J., Littin, J. and Witten, I.H., 1997. Applications of machine learning in information retrieval. *Annual Review of Information Science* 34, pp. 341-384.
- Dahlback, N., Joensuu, A. and Ahrenberg, L., 1993. Wizard of Oz Studies: Why And How. In Proceedings ACM International Workshop on Intelligent User Interfaces, ACM Press, New York.
- Dautenhahn, K., 1998. The art of designing socially intelligent agents. *Science, fiction and the human in the loop*. *Applied Artificial Intelligence*, 12, 8-9, pp. 573-617.
- de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A. and De Carolis, B., 2006. User modeling and adaptation in health promotion dialogs with an animated character. *Journal of Biomedical Informatics*, Special Issue on 'Dialog systems for health communications'. 39, 5, pp. 514-531.
- de Rosis F, Batliner, A., Novielli N. and Steidl S., 2007. 'You are sooo cool Valentina! Recognizing social attitude in speech-based dialogues with an ECA. In A.Paiva, R.Prada and R.Picard (Eds.). Proceedings of ACII 2007. Springer LNCS 4738.
- Devillers, L. and Vidrascu, L., 2006. Real-life emotion detection with lexical and paralinguistic cues on human-human call center dialogues, in: Proceedings of INTERSPEECH.
- Di Eugenio, B., 2000. On the usage of Kappa to evaluate agreement on coding tasks, in: LREC2000: Proceedings of the Second International Conference on Language Resources and Evaluation, pp. 441-444.

- Di Eugenio, B. and Glass, M., 2004. The Kappa statistics, a second look, in: *Computational Linguistics*, vol. 30, pp. 95-101.
- Ekman, P., 1999. Basic Emotions. In T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion*. Sussex, U.K.: John Wiley & Sons, Ltd , pp. 45-60.
- Fellbaum, C., 1998. *WordNet. An Electronical Lexical Database*. The MIT Press.
- Gill, A.J. and Oberlander, J., 2002. Taking care of the linguistic features of extraversion, in: *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pp. 363-368.
- Goetz, J., Kiesler, S. and Powers, A., 2003. Matching robot appearance and behavior to task, to improve human-robot cooperation, in: *Proceedings of ROMAN*, pp. 55-60.
- Höök, K., Isbister, K. and Laksolathi, J., 2005. Sensual evaluation instrument, in: *Proceedings of CHI Workshop on Innovative Approaches to Affective Evaluation*.
- Landauer, T.K. and Dumais, S.T., 1997. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, pp. 211-240.
- Lee, C.M., Narayanan, S. S., and Pieraccini, R., 2002. Combining acoustic and language information for emotion recognition, in: *Proceedings of the 7th International Conference on Spoken Language Processing*, pp. 873-876.
- Litman, D., Forbes-Riley, K., Silliman, S., 2003. Towards emotion prediction in spoken tutoring dialogues, in: *Proceedings of HLT/NAACL*, pp. 52-54.
- Liu, H. and Maes, P., 2004. What would they think? A computational model of attitudes. *Proceedings of the International Conference on Intelligent User Interfaces*, pp. 38-45.
- Mazzotta, I., Novielli, N., Silvestri, V. and de Rosis, F., 2007. O Francesca, ma che sei grulla? Emotions and irony in persuasion dialogues. *Proceedings of the 10th Conference of AI*IA - Special Track on 'AI for Expressive Media'*. AI*IA 2007: Artificial Intelligence and Human-Oriented Computing, Rome, 602-613.
- Nass, C., Isbister, K. and Lee, E-J, 2000. Truth is beauty: Researching embodied conversational agents, in: Cassell, J. Sullivan, S. Prevost and E. Churchill (Eds.), *Embodied Conversational Agents*, The MIT Press.
- Neviarouskaya, A., Prendinger, H. and Ishizuka, M., 2007. Analysis of affect expressed through the evolving language of online communication. In the *Proceedings of the International Conference on Intelligent User Interfaces*. ACM Press, pp. 278-281.
- Niederhoffer, K.G. and Pennebaker, J.W., 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*. 21, 4, pp. 337-360.
- Oviatt, S.L., Cohen, P.R. and Wang, M., 1994. Towards interface design for human language technology: modality and structure as determinants of linguistic complexity. *Speech Communications*. 15, pp. 283-300.

- Oviatt, S.L. and Adams, B., 2000. Designing and evaluating conversational interfaces with animated characters, in: J. Cassell, J. Sullivan, S. Prevost and E. Churchill (Eds.), *Embodied Conversational Agents*, The MIT Press.
- Pang, B., Lee, L. and Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques, in: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 76-86.
- Picard, R. and Bryant Daily, S., 2005. Evaluating affective interactions: alternatives to asking what users feel, in: *Proceedings of CHI Workshop on Evaluating Affective Interfaces*.
- Polhemus, L., Shih, L-F and Swan, K., 2001. Virtual interactivity: the representation of social presence in an on line discussion. Annual Meeting of the American Educational Research Association.
- Rettie, R., 2003. Connectedness, awareness and social presence, in: *Proceedings of PRESENCE*, online proceedings.
- Rosenfeld, R., 2000. Two decades of statistical language modeling: where do we go from here?, in: *Proceedings of the IEEE*, 88, 8, pp. 1270-1278.
- Schneiderman, B. and Maes, P., 1997. Direct manipulation vs. interface agents. *Interactions*, 4, 6, pp. 42-61.
- Scherer, K.R., Wranik, T., Sangsue, J., Tran, V., and Scherer, U., 2004. Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction pattern. *Social Science Information*. 43, 4, pp. 499-570.
- Sidner, C. and Lee, C., 2003. An architecture for engagement in collaborative conversations between a robot and a human. MERL Technical Report, TR2003-12.
- Swan, K., 2002. Immediacy, social presence and asynchronous discussion, in: J. Bourne and J.C.Moore (Eds.): *Elements of quality online education*. Vol3, Nedham, MA. Sloan Center For Online Education.
- Yu, C., Aoki, P.M. and Woodruff, A., 2004. Detecting user engagement in everyday conversations, in: *Proceedings of 8th International Conference on Spoken Language processing*, pp. 1329-1332.
- Walton, D., 2006. Examination dialogue: an argumentation framework for critically questioning an expert opinion. *Journal of Pragmatics*, 38,745-777
- Whitelaw, C., Argamon, S. and Garg, N., 2005. Using appraisal taxonomies for sentiment analysis. *Proceedings of the First Computational Systemic Functional Grammar Conference*, Sydney.
- Wilson, T., Wiebe, J and Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment *analysis*. *Proceedings of HLT-EMNLP Vancouver*.
- Zhang, L., Barnden, J.A., Hendley, R.J. and Wallington, A.M., 2006. Exploitation in affect detection in open-ended improvisational test, in: *Proceedings of the Workshop on Sentiment and Subjectivity in Text, ACL*, pp. 47-54.

Zimmerman, J., Ayoob, E., Forlizzi, J. and McQuaid, M., 2005. Putting a face on embodied interface agents, in: Proceedings of the Conference on Designing Pleasurable Products and Interfaces, Eindhoven Technical University Press, pp. 233-248.

Zweig H.H. and Campbell, G., 1993. Receiver Operating Characteristic (ROC) plots: a fundamental evolution tool in medicine. *Clinical Chemistry*, 39, pp. 561-77.

FIGURE CAPTIONS

Figure 1. Architecture of our Wizard of Oz tool

Figure 2. The online evaluation questionnaire

Figure 3. ROC curves for the Bayesian classifier (written input)

Table 1. Our markup language for signs of social attitude

Sign with definition	Example
<p>Friendly self-introduction</p> <p>The subjects introduce themselves with a friendly attitude (e.g. by giving her name or by explaining the reasons why they are participating in the dialogue)</p>	<p>Oz: Hi. My name is Valentina. I'm here to suggest you how to improve your diet. S: Hi, my name is Isa and I'm curious to get some information about healthy eating</p>
<p>Colloquial style</p> <p>The subject employs a current language, dialectal forms, proverbs etc</p>	<p>Oz: Are you attracted by sweets? S: I'm crazy for them.</p>
<p>Talks about self</p> <p>The subject provides more personal information about self than requested by the agent</p> <p><i>Personal questions to the agent.</i></p> <p>The subject tries to know something about the agent preferences, lifestyle etc, or to give it suggestions.</p>	<p>Oz: Do you like sweets? Do you ever stop in front of the display window of a beautiful bakery? S: Very much! I'm greedy! Oz: What did you eat at lunch? S: Meat-stuffed peppers. How about you?</p>
<p>Humor and irony</p> <p>The subjects make any kind of verbal joke in their move</p>	<p>Oz: I know we risk to enter into private issues. But did you ever try to ask yourself which are the reasons of your eating habits? S: Unbridled life, with light aversion towards healthy food.</p>
<p>Positive or negative comments</p> <p>The subjects comment the agent behavior in the dialogue: its experience, its degree of domain knowledge, the length of its moves etc.</p>	<p>Oz: I'm sorry, I'm not much an expert in this domain. S: OK: but try to get more informed, right? Oz: Good bye. S: What are you doing? You leave me this way? You are rude!!</p>
<p>Friendly farewell</p> <p>The subject uses a friendly farewell form or asks to carry-on the dialogue.</p>	<p>Oz: Goodbye. It was really pleasant to interact with you. Come back when you wish. S: But I would like to chat a bit more with you.</p>

Table 2. Frequency of signs of social attitude and inter-rater agreement

Signs		WRITTEN INPUT			SPOKEN INPUT		
		frequency	agreement	kappa	frequency	agreement	kappa
Friendly self-introduction (fsi)		2%	0,98	0,87	2%	0,99	0,86
Colloquial style (cstyle)		9%	0,89	0,70	11%	0,91	0,65
Talks about self (talks)		19%	0,73	0,64	21%	0,87	0,59
Questions about the agent (qagt)		12%	0,70	0,56	7%	0,92	0,53
Comments	Positive (pcomm)	4%	0,82	0,42	7%	0,90	0,41
	Negative (ncomm)	5%	0,86		5%	0,94	
Friendly farewell (ffwell)		4%	0,93	0,65	4%	0,98	0,76

Table 3. multiple regression model for the percentage of 'social' moves in subject's dialogues

VARIABLE*	COEFFICIENT	STANDARD ERROR	T	ONE-SIDED P
Intercept	.08	.08	1.04	.15
Interaction mode	-.07	.05	1.29	.10
Gender	.04	.05	.81	.21
Background	.12	.05	2.21	.02
Number of Moves	.004	.002	1.84	.04
Average move length	.004	.0006	6.23	.0000
<i>R-squared = 0.63; st. error of estimate = .18</i>				
<i>The following dummy variables were introduced: gender was coded as M=0, F=1; interaction mode was coded as written=0, spoken=1; background was coded as CS=0, H=1.</i>				

Table 4. signs and linguistic categories, with some examples

Signs	Linguistic categories	Examples
FRIENDLY	Greetings	Good morning, nice to meet you, ...
SELF-INTRODUCTION (fsi)	Self introduction Ciao	My name is, I am, ... Ciao, ciao, ...
Colloquial style (cstyle)	Paralanguage Terms From Spoken Language Dialectal and Colored Forms Proverbs and Idiomatic Expressions Diminutive or Expressive Forms	!, hurrah, .. Ok, siiiii (yees), chiacchierare (to chat), la mia passione (my passion)... vabbé, a me mi, puccia, espressino... mens sana in corpore sano, carta bianca (carte blanche)... Ciccione (fatty), piccolina, ...
Talks about self (talks)	First person pronouns First person auxiliary verbs First person knowledge verbs First person attitude verbs First person ability verbs First person liking or desiring verbs <i>First person domain verbs</i>	I, my, to me, for me... I have, I am, ... I know, believe, ... I try, do, tend to, ... I can, succeed, I like, would like, prefer, ... want, care,... <i>I drink, eat,</i>
Questions about the agent (qagt)	Second person pronouns Second person auxiliary verbs Second person knowledge verbs Second person attitude verbs Second person ability verbs Second person liking or desiring verbs <i>Second person domain verbs</i>	You, your, to you,.. You have, are, ... You know, believe, ... You try, do, tend to, ... You can, succeed, You like, would like, prefer, ... want, care,... <i>You drink, eat,</i>
Positive or Negative comments (poscom / negcom)	Generic comments Expressions of agreement or disagreement Message evaluation Evaluation of agent's politeness Evaluation of agent's competence Remark about agent's repetitivity Evaluation of agent's understanding ability	You scare me, your depress me, stop please, ... I agree, you're right ... but, what's bad?, I don't agree, ... It's too much, it's not enough, ... You are kind, rude, ... You (don't) know, you are (not) able to / narrow-minded, ... You repeat the same thing, you already told it, ... You (don't) understand, ...
Friendly farewell (ffwell)	Expressions of farewell Thanking Ciao	See you, bye... Thanks, thank you, ... Ciao, ciao, ...
All examples, except those referring to 'colloquial style', are translated from Italian.		

Table 5. Accuracy of Bayesian classification in recognizing the various signs (written input)

	SENSITIVITY	SPECIFICITY	ACCURACY
fsi	1.00	.96	.96
cstyle	.89	.77	.79
talks	.81	.88	.85
qagt	.92	.86	.87
poscom	.80	.95	.94
negcom	.73	.94	.92
ffwell	.93	.96	.96

Table 6. Results of Bayesian classification: confusion matrix (written input)

	FSI	CSTYLE	TALKS	QAGT	POSCOM	NEGCOM	FFWELL	NEUTRAL
fsi	1.00	0.14	0.14	0.00	0.00	0.00	0.86	0.00
cstyle	0.00	0.89	0.04	0.15	0.22	0.07	0.11	0.04
talks	0.00	0.16	0.81	0.03	0.03	0.10	0.01	0.14
qagt	0.00	0.13	0.08	0.92	0.00	0.04	0.00	0.04
poscom	0.00	0.35	0.25	0.25	0.80	0.00	0.20	0.00
negcom	0.00	0.20	0.27	0.40	0.00	0.73	0.00	0.20
ffwell	0.55	0.67	0.00	0.40	0.00	0.00	0.93	0.07

Table 7. Accuracy of Bayesian classification in recognizing the various signs (spoken input)

	SENSITIVITY	SPECIFICITY	ACCURACY
fsi	1.00	.93	.93
cstyle	.79	.69	.70
talks	.83	.67	.74
qagt	.85	.78	.79
poscom	.68	.78	.77
negcom	.68	.73	.73
ffwell	1.00	.93	.94

Table 8. Confusion matrix after revision of categories and parameters (spoken input)

	fsi	cstyle	talks	qagt	poscom	negcom	ffwell	neutral
fsi	1.00	0.28	0.44	0.33	0.11	0.00	1.00	0.00
cstyle	0.00	0.79	0.34	0.17	0.21	0.24	0.00	0.03
talks	0.01	0.29	0.83	0.17	0.19	0.47	0.02	0.05
qagt	0.04	0.17	0.37	0.85	0.22	0.11	0.02	0.04
poscom	0.00	0.30	0.28	0.38	0.68	0.06	0.26	0.06
negcom	0.00	0.44	0.40	0.28	0.08	0.68	0.00	0.08
ffwell	0.67	0.57	0.17	0.17	0.50	0.03	1.00	0.00