# A Model for Synthesizing a Combined Verbal and Nonverbal Behavior Based on Personality Traits in Human-Robot Interaction

Amir Aly, Adriana Tapus

# A Model for Synthesizing a Combined Verbal and NonVerbal Behavior Based on Personality Traits in Human-Robot Interaction

Amir Aly and Adriana Tapus
Cognitive Robotics Lab
ENSTA-ParisTech
828 Blvd Marechaux, Palaiseau, France
Email:{amir.aly; adriana.tapus}@ensta-paristech.fr

Abstract—Robots are more and more present in our daily life; they have to move into human-centered environments, to interact with humans, and to obey some social rules so as to produce an appropriate social behavior in accordance with human profile (i.e., personality, state of mood, and preferences). Recent research studies discussed the effect of personality traits on the verbal and nonverbal production, which plays a major role in transferring and understanding messages in a social interaction between a human and a robot. The characteristics of the generated gestures (e.g., amplitude, direction, rate, and speed) during the nonverbal communication can differ according to the personality trait, which, similarly, influences the verbal content of the human speech in terms of verbosity, repetitions, etc. Therefore, our research tries to map human verbal behavior to a corresponding combined robot verbal-nonverbal behavior based on the personality dimensions of the interacting human. The system estimates - first - human personality traits through a psycholinguistic analysis of the spoken language, then it uses PERSONAGE natural language generator that tries to generate a corresponding verbal language to the estimated personality traits. Gestures are generated by using BEAT toolkit, which performs a linguistic and contextual analysis of the generated language relying on rules derived from extensive research into human conversational behavior. We explored the human-robot personality matching aspect and the differences of the adapted combined robot behavior (gesture and speech) over the adapted speech only robot behavior in interaction. Our model validated that the participants preferred more to interact with a robot that has a similar personality. Besides, it validated that the adapted robot behavior expressed through combined gestures and speech, is more engaging and attractive than the adapted robot behavior expressed only through speech. Our experiments were validated using Nao robot.

## I. Introduction

Creating a socially-intelligent robot able to interact with humans in a natural manner and to synthesize appropriately comprehensible multimodal behaviors in a wide range of interaction contexts, is a highly complicated task. This requires a high level of multimodal perception, so that the robot should understand the internal states, intentions, and personality dimensions of the human in order to be capable of generating an appropriate verbal and nonverbal combined behavior.

The related literature reveals hard efforts aiming to support the natural human-robot conversational interaction. The authors in [1], tried to create a limited verbal natural language interface to access information in a database. An interesting theoretical study on the Natural Language (NL) was discussed in [2], in which the authors tried to study the effect of using natural language interaction of rich functionality (e.g. paraphrasing, correcting misconceptions, etc.) on the effective use of expert systems. Another interesting theoretical study was discussed in [3] and [4], where the authors focused on the field of user modeling (i.e. understanding user beliefs, goals and plans) in artificial intelligence dialog systems, and illustrated the importance of such modeling on interaction. Later on, some research studies tried to illustrate how the dialogue systems that are adapted to the user's model (including the ability to explicitly and dynamically change the aspects of the relationship with the interacting human through the use of social talks, in the same way as humans behave) will be more believable [5],[6],[7],[8].

Some efforts were driven towards generating synchronized verbal and nonverbal behaviors as discussed in [9]. The authors presented a model that is capable of synchronizing expressive gestures with speech. The model was implemented on Honda humanoid robot and was able to generate a full range of gesture types, such as emblems, iconic and metaphoric gestures, deictic pointing and beat gestures. Similarly, virtual agents had received much attention concerning generating expressive behaviors. The authors in [10] tried to simulate the natural speech-gestures production model that humans have on the 3D agent MAX. They proposed an architecture for generating synchronized speech and gestures in a free and spontaneous way. For example, it is sufficient to support the system with some preliminary information about an object to describe and the system will generate itself an expressive combined verbal and nonverbal behavior exactly as humans do. Another interesting approach was discussed in [11], [12]. The authors developed the embodied conversational agent GRETA. It can express its emotional states and intentions through verbal and nonverbal behaviors, and be socially aware. It can endorse different roles; e.g. be a dialog companion, a storyteller, a virtual tutor, or even an actor in different games. Despite the rich literature of generating expressive behaviors with humanoid robots and 3D agents, and to the best of our knowledge, no research work discussed the importance of generating combined verbal and nonverbal behaviors based on the personality traits of the interacting human.

Personality is a key determinant in human social interactions. In the literature, there are different models of personality (e.g., Big5 [13], Eysenck Model of Personality (PEN) [14], Meyers-Briggs [15]). The personality has a long term effect on the generated behavior, which gives more reliability to the personality dimensions for characterizing the generated verbal and nonverbal behavior, to the contrary of other short-term characteristics like estimating human emotions through prosodic features. In HRI and HCI, a direct relationship between personality and behavior has long been recognized [16], [17], [18], [19]. In the context of human modeling and adapting the dialog of a machine (i.e. a humanoid robot or a computer) to the personality of the interacting human, the authors in [20], [21], [18] proved empirically that the human interacting with a dialog machine will spend more time on the assigned task if the generated behavior matches with human personality, which validates the similarity attraction principle stating that individuals are more attracted by others with the same personality traits.

Based on these findings, we posit that the personality is a key factor in human-robot interactions (HRI). In this research, we try to develop an adapted customized verbal-nonverbal robot behavior based on the personality dimensions of the interacting human. The work described in this paper validates that individuals prefer more to interact with a robot that has the same personality with theirs and that a combined (gesture and speech) robot behavior is more engaging than a speech only robot behavior.

The rest of the paper is structured as following: Section II presents a general overview of the architecture of the system, Section III describes the nonverbal behavior's knowledge base extension, Section IV illustrates how we realized the synchronized verbal and nonverbal behaviors on the robot, Section V illustrates the design, the hypotheses, and the scenario of interaction, Section VI provides a description of the experimental results, and finally, Section VII concludes the paper.

## II. SYSTEM ARCHITECTURE

Our system is a coordination between many subsystems: (1) Dragon Naturally Speaking toolkit that translates the spoken language of the interacting human into a text; (2) Personality Recognizer that tries to estimate the personality traits of the interacting human [22]; (3) PERSONAGE natural language generator that adapts the generated text to the personality dimensions of the interacting human [23]; (4) BEAT toolkit that translates the generated text into gestures[24]; (5) Nao robot as the test-bed platform. An overview of our system architecture is presented in Figure 1.

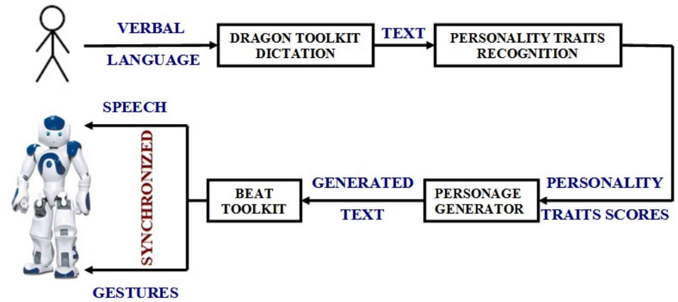


Fig. 1. General Overview of the System Architecture

### A. Personality Recognizer

Personality markers in language had received lot of interest from psycholinguistic studies. The authors in [25],[26], and [27] described how linguistic features linked the extraversion and the introversion traits to speech production. They stated that extraverts talked more, louder, and more repetitively with less pauses and hesitations, with respect to the introverts. Moreover, extraverts have higher speech rates, a higher verbal output, and a less formal language, while introverts use a broader vocabulary. On the other hand, the authors in [28], stated that extraverts use more positive emotion words, and show more agreements and compliments than introverts.

A general approach for characterizing the majority of personality traits was discussed in [28], in which the authors used the Linguistic Inquiry and Word Count (LIWC) tool to count word categories of 2479 essays (containing 1.9 million words) written by different persons covering the five personality traits described in Big Five Framework (Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). This dictionary enabled them to state general relationships and characteristics of the five personality traits, such as: Conscientious people avoid negations, negative emotion words, and words reflecting discrepancies (e.g., should and would). Similarly, the authors in [29] created a spoken data corpus (97468 words and 15269 utterances) beside their transcripts, covering different personality traits. This corpus was sub-divided into several word categories using the LIWC tool.

The findings of the previous two data corpora were the basic body of the research conducted by the authors in [22]. They created a huge database including LIWC psycholinguistic features, such as: Anger words (e.g. hate), Metaphysical issues (e.g. God), Family members (e.g. mom, brother), etc., in addition to other features from the MRC psycholinguistic database [30], such as: Frequency of use (e.g. Low: duly, nudity, High: he, the), Concreteness (e.g. Low: patience, High: ship), etc., in addition to the Utterance type features, such as: Command (e.g. must, have to), Prompt (e.g. Yeah, Ok), Question, and Assertion (which is any utterance out of the previous categories). The relationship between the utterance type features and the personality traits was discussed in [31],[32], in which for example, extraverts are more assertive when writing emails. Afterwards, the system was trained on the previously stated data corpora using the support vector machines (SVM) algorithm, and was cross validated so as to approve its performance.

## B. PERSONAGE Generator

PERSONAGE is a natural language generator that can model findings from psychological studies to express various personality traits through language. The architecture of the generator is illustrated in Figure (2), which is based on the traditional pipelined natural language generation (NLG) architecture [33]. The input consists of personality traits' scores besides a selection of restaurants (recommendation and/or comparison) in New York City, with associated scalar values representing the ratings of six attributes: food quality, cuisine, service, location, price and atmosphere (PERSONAGE is previously supported by a database containing the six attributes ratings of more than 700 restaurants in New York City, the data is collected from real surveys and opinions of people that visited these restaurants). In addition to other generation parameters $\in [0, 1]$, like the verbosity parameter which is set to 1, in order to maximize the wordy content in the utterance. The content planner is responsible for selecting and structuring (in a tree format) the information to be processed by the sentence planner in terms of the values of some parameters such as the verbosity, polarity, repetitions (i.e. the content planer decides what to say).

The sentence planner deals with phrasing the information structured by the content planner. It searches in the generation dictionary for the set of syntactic elementary structures stored for each proposition in the content plan (e.g. if the content planner structured a recommendation, the sentence planner would precise the syntactic parts of the recommendation like: verb, noun, etc.). Afterwards, it aggregates the obtained syntactic templates in order to produce the utterance's full syntactic structure [34].

The pragmatic marker insertion process in the sentence planner modifies the aggregated syntactic structure in order to produce various pragmatic effects like the hedge *you know*, the question tags, etc. The lexical choice process selects the most appropriate lexeme (from many different lexemes expressed by PERSONAGE) for each content word in terms of the lexeme's length, frequency of use, and strength [35],[36]. Last but not least, the realization process, which follows the sentence planner, converts the final syntactic structure into a string by applying surface grammatical rules, such as morphological inflection and function word insertion [37].

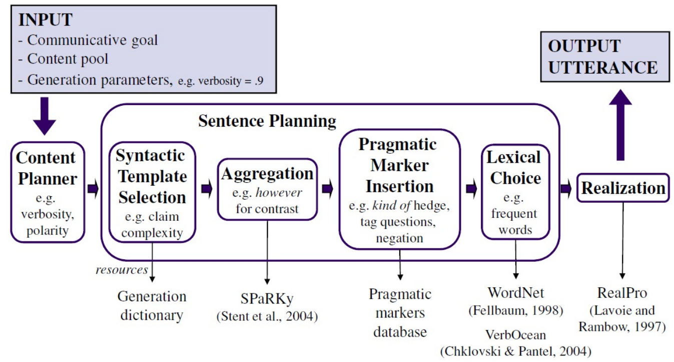


Fig. 2. PERSONAGE Architecture [23]

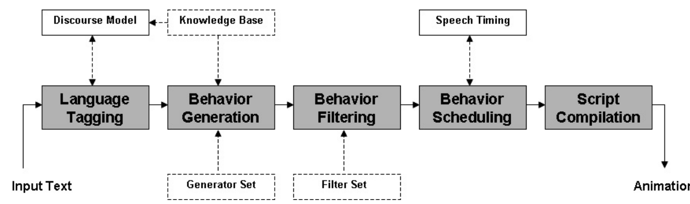


Fig. 3. Beat Architecture [24]

## C. BEAT Toolkit

BEAT is the Behavior Expression Animation Toolkit that takes a text as input and generates a corresponding synchronized set of gestures. It uses linguistic and contextual information contained in the text to control body and face gestures, besides the intonation of the voice. The mapping from text to facial, intonational, and bodily gestures is contained in a set of rules derived from the state of the art in nonverbal conversational behavior research [24].

BEAT pipeline is composed of different XML- based modules as illustrated in Figure (3). The language tagging module receives an XML tagged text generated from PERSONAGE, and converts it into a parse tree with different discourse annotations (e.g. theme and rheme). The behavior generation module uses the output tags of the language module and suggests all possible gestures, then the behavior filtering module selects the most appropriate set of gestures (using the gestures conflict and priority threshold filters). The user-definable data structures, such as: the generator and filter sets (indicated in dotted lines) provide the generation and filtering rules and conditions for the behavior generation and selection processes. Meanwhile, the knowledge base adds some contextual information and definitions that are important in generating relevant and precise non verbal behaviors, such as: *Type* which attributes features with their values to different object types (e.g., the object "Home", which belongs to the class "Place" with type features attributes as "House, Apartment"), *Instance*, which describes specific cases of recognizable objects (e.g., the "Spiral" shape could be considered as a shape instance of the object "Stairs"), *Scene*, which groups all instances of the same environment into scenes, and *Gesture*, which specifies different kinds of gestures and their proposed trajectories and hand shapes. The behavior scheduling module converts the input XML tree into a set of synchronized speech and gestures. It includes a TTS (text-to-speech) engine that calculates the duration of words and phonemes, which helps in constructing an animation schedule for the aligned gestures with words. The script compilation module compiles the animation script into some executive instructions that can be used in animating a 3D agent or a humanoid robot.

## III. NONVERBAL BEHAVIOR'S KNOWLEDGE BASE EXTENSION

The purpose of the extension we performed on BEAT toolkit is to add necessary information about the generated text by PERSONAGE comparing (and/or recommending) between different restaurants in New York City. The object type "Restaurant" is defined as an object in the class "Place" with some information about the location, price category, and cuisine, which have been used in the interaction scenarios. Some instances were also added to the knowledge base describing some attached places to the object "Restaurant", such as: "Basement", and "Dining Room", in terms of the size, lightening, and painting. The new added scenes to the knowledge base define the name of restaurants containing the previously defined instances. The precised gestures' characteristics in the knowledge base concern different types of iconic gestures (beside deictic gestures) with their hand shapes and arm trajectories. Some new added linguistic keywords were aligned to specific iconic gestures with their geometrical hand/arm shapes' characteristics, such as: the adjective "narrow" was aligned to the hand shape "Hands in Front" and the arm trajectory "Span" in order to refer to a small span separating between the two hands which match the meaning of the adjective "narrow".

## IV. MODELING THE SYNCHRONIZED VERBAL AND NON VERBAL BEHAVIORS ON THE ROBOT

Despite the fact that BEAT toolkit was built as a customizable gestures generator, so that some other categories of gestures can be added to the generation system, we found that the already built in gestures are sufficient for the short verbal context generated by PERSONAGE. Therefore, in this research we are interested in four kinds of gestures: Iconic, Posture Shift, Metaphoric, and Gaze gestures. The animation script (generated by BEAT toolkit) described in Figure (4), indicates the proposed synchrony between the verbal content and the corresponding allocated gestures of the following sentence: The first restaurant was calm and not far from downtown but expensive though. The second restaurant had a narrow dining room but also had a better quality and was little cheaper. The system divides the sentence into chunks, where each chunk contains a group of words with a specific allocated gesture. The symbol "WI" indicates the index of words (31 words in total), while the symbol "SRT" defines the estimated duration of each group of words with the allocated gesture. The animation script reveals also that the adjective word "narrow" was attributed to an iconic gesture where the two hands will be used to depict the gesture "Gesture Both", which proves the importance of customizing the knowledge base in order to generate the most appropriate nonverbal behavior.

Metaphoric gestures (which are not present in the animation script in Figure 4) are used frequently in order to represent the narrated speech but not in a physical way like the iconic gestures. They could take the form of a general arm/hand shaking or even a specific shape, like when we want to express time sequences, we use the word "after" associated with a specific arm/hand motion symbolizing this idea. Therefore, this word (and other similar new words) is added and allocated in the knowledge base to the corresponding specific arm/hand motion trajectory similarly to iconic gestures. On the other hand, the generation of general metaphoric gestures doesn't follow a specific linguistic rule, which makes it a virtual generation of gestures. Our approach associates the generation of general metaphoric gestures to some prosodic rules as discussed in [38],[39], so as to integrate the para-verbal modality into the generation of a nonverbal behavior. Consequently, for an introverted speaker who doesn't speak a lot, he will have a corresponding limited pitch contour which will lead to a corresponding limited set of generated gestures, and vise versa for the extraverted individuals.

The mapping of the gaze, posture-shift, iconic, and specific-shape metaphoric gestures to the robot from the animation script necessitates that the robot processes each line of the script indicating the duration of each chunk containing a synchronized verbal content with a nonverbal behavior. Kendon in [40] defined the gesture phrases as the units of gestural movement that contain one or more subsequent movement phases, which are: preparation, stroke, and retraction beside some intermediate holds. The problem that may appear when modeling a combined verbal and nonverbal behavior on a robot (in case of iconic and specific-shape metaphoric gestures), is the required high temporal synchronization between the stroke (the expressive phase) and the affiliate (the affiliated word or sub-phrase) in order to express an idea accurately. The time estimation indicated in the animation script reveals the estimated time for the stroke phase (in case of a generated iconic gesture), consequently an additional time estimation for the preparation phase should be assumed so that the hands/arms leave their initial position and get ready for the stroke phase. Therefore, the gesture stroke is set to precede the affiliate's onset by a given offset (one syllable's approximate duration of 0.3 s).

```xml
<?xml version="1.0" encoding="UTF-8"?>
- <AnimationScript HEARER="USER" SPEAKER="AGENT">
    <START SRT="0.0" WI="0" SPEECH=" The first restaurant was calm and not far from downtown but expensive
     though.The second restaurant had a narrow dining room but also had a better quality and was little cheaper.
     " ACTION="SPEAK" AID="A527"/>
    <START SRT="0.0" WI="0" ACTION="GAZE" AID="A535" PRIORITY="1" DIRECTION="AWAY_FROM_HEARER"/>
    <START SRT="0.0" WI="0" ACTION="POSTURESHIFT" AID="A539" ENERGY="LOW" BODYPART="BOTH"/>
    <STOP SRT="0.6" WI="1" ACTION="POSTURESHIFT" AID="A539" ENERGY="LOW" BODYPART="BOTH"/>
    <STOP SRT="2.5" WI="3" ACTION="GAZE" AID="A535" PRIORITY="1" DIRECTION="AWAY_FROM_HEARER"/>
    <START SRT="2.5" WI="3" ACTION="GAZE" AID="A552" PRIORITY="5" DIRECTION="TOWARDS_HEARER" FOCUS="ANY"/>
    <STOP SRT="9.1" WI="18" ACTION="GAZE" AID="A552" PRIORITY="5" DIRECTION="TOWARDS_HEARER" FOCUS="ANY"/>
    <START SRT="9.1" WI="18" ACTION="GAZE" AID="A552" PRIORITY="5" DIRECTION="TOWARDS_HEARER" FOCUS="ANY"/>
    <STOP SRT="9.8" WI="19" ACTION="GESTURE_BOTH" AID="A563" PRIORITY="20" TRAJECTORY="SPAN" SHAPE=
     "HANDS_IN_FRONT"/>
    <START SRT="9.8" WI="19" ACTION="GAZE" AID="A552" PRIORITY="5" DIRECTION="TOWARDS_HEARER" FOCUS="ANY"/>
    <STOP SRT="15.6" WI="30" ACTION="GAZE" AID="A552" PRIORITY="5" DIRECTION="TOWARDS_HEARER" FOCUS="ANY"/>
  </AnimationScript>
```


Fig. 4. Synchronized Verbal and Non Verbal XML Animation Script

## V. EXPERIMENTAL SETUP

We begin this section by introducing the robot system that was used in our experiments and continue with the overview of the experiments we conducted.

### A. Robot test-bed

The experimental test-bed used in this study is the humanoid Nao robot developed by Aldebaran Robotics[1]. Nao is a 25 degrees of freedom robot, equipped with an inertial sensor, two cameras, eyes eight full-color RGB LEDs, and many other sensors, including a sonar which allows it to comprehend its environment with stability and precision.

### B. Hypotheses

The presented work aimed to test the following hypotheses:

Hypothesis 1: The robot behavior that matches the user's personality expressed through speech and gestures will be preferred by the user. Hypothesis 2: The robot personality expressed through adapted combined speech and gestures will be perceived more expressive by the user than the robot personality expressed only through adapted speech.

### C. Experimental Design

In order to test and validate the first hypothesis, the user was exposed to 2 robot personalities:

- the robot uses introverted cues (expressed through gestures and speech) to communicate with the user.
- the robot uses extraverted cues (expressed through gestures and speech) to communicate with the user.

Similarly, in order to validate the second hypothesis, the users tested 2 different conditions:

- the robot communicates with the user only through speech (the robot-user personality matches). We call it: adapted speech only robot behavior.
- the robot communicates both through gestures and speech with the user (the robot-user personality matches), We call it: adapted combined robot behavior.

All the previous four conditions were randomly ordered during the experiments. For the second hypothesis, we excluded the condition of interaction through gestures only, as it doesn't fit in the normal context of non-mute human-human interaction. Before the experiments, each participant was asked to complete two questionnaires: one for determining personal details such as gender, age, occupation, and educational background, and another for establishing the subject's personality traits based on the Big 5 Inventory Test [41]. For our experiments, we focused only on the extraversion factor that indicates the level of sociability of an individual. An extraverted individual tends to be sociable, friendly, fun loving, and talkative, while an introverted individual tends to be reserved, inhibited, and quiet (Figure 5).

[1]http://www.aldebaran-robotics.com/

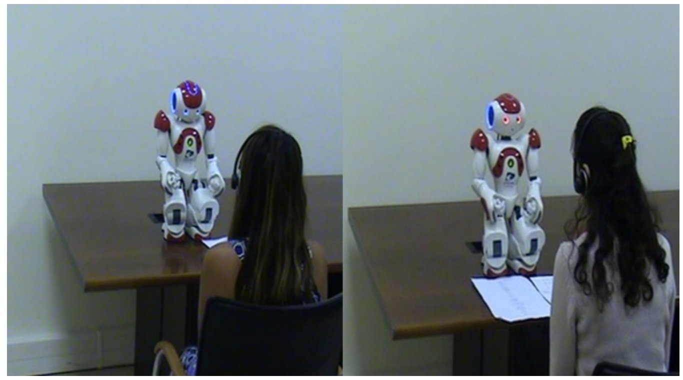

Fig. 5. Introverted and Extraverted robot conditions (In the introverted condition, the robot's head was looking down with a low gesture rate. Meanwhile in the extraverted condition, the robot's head was looking up with a high gesture rate).

The theme of the interaction is restaurants in New York city (USA). The robot has a list of restaurants from New York city and its role is to give advice regarding the food quality, the service, and the price of the restaurants.

Our interaction scenario is as following:

- The robot introduces itself and asks the participant to say some things he/she knows about New York city. This first step is used by the robot to automatically identify the user's personality based on the linguistic cues (the Personality Recognizer tool is used).
- Afterwards, the robot has a list of restaurants and asks the participant to choose a particular restaurant so as to find out more details about it.
- The robot waits for the participant's input and produces an appropriate utterance and gesture based on the personality traits (PERSONAGE generator and BEAT toolkit are employed).
- The participant can ask details about other restaurants.
- The interaction ends when the user doesn't want to know more information about other restaurants.

At the end of each experiment, the experimenter presented a short debriefing. The average time of a single interaction in a given condition was varying between 3 and 4 minutes. This variation depends on the fact that some participants asked more questions than the others during the free interaction experiments, without any relation to the personality condition.

The system evaluation was performed based on user introspection (questionnaires). After each experiment, the participant completed one questionnaire designed to evaluate the impression of the robot personality, the interaction with the robot, the robot speech and gestures synchronization and matching, etc. All questions (24 question in total) were presented on a 7-point Likert scale.

## VI. EXPERIMENTAL RESULTS

The subject pool for this experiment consisted of 21 participants (14 male, 7 female; 12 introverted and 9 extraverted). Approximately 47% were 21-25 years old and 53% were 26-30 years old. The recruited participants were ENSTA-ParisTech undergraduate and graduate students, diverse in undergraduate major. We found that there is a reasonable correlation between the measured online and offline (Big 5 Inventory Test [41]) personality traits of the participants in 14 results. For the 7 uncorrelated results, 4 results were considered as introverted in the offline test and became extraverted in the online test, and vise versa for the remaining uncorrelated 3 results. However, the online test is believed to be more trusted than the offline test, because people may show different aspects of their personality in an imprecise manner through the offline test.

In order to test the first hypothesis all the participants were exposed to two robot conditions: introverted robot and extraverted robot. The robot personality was expressed through speech and gestures as discussed in Section II. In the introverted robot condition, the robot gestures were narrow, slow and at a low rate. Contrarily, in the extraverted robot condition, the robot gestures were broad, quick, and at a high rate (Figure 5). The speech content is also based on personality traits; the robot gives more details in the extraverted condition than in the introverted condition. After each condition, the participants were asked to answer the following question: "Do you find a match between the robot personality (expressed through speech and gestures) and your personality?".

Our ANOVA analyses showed that the extraverted individuals perceived the extraverted robot as significantly more close to their personality than the introverted robot ($F[1,17] = 40.5$, $p = 0.0$). A similar tendency was observed for the introverted individuals who preferred the introverted robot to the extraverted robot ($F[1,23] = 7.76$, $p = 0.0108$) (Figure 6). All participants (introverted and extraverted together) considered that the robot speech and gestures were semantically matched (content), significantly more in the extraverted condition than in the introverted condition ($F[1,41] = 9.29$, $p = 0.0041$). However, when the personality trait is included in the analysis, this aspect is significant only for the extraverted individuals ($F[1,17] = 6.87$, $p = 0.0185$).

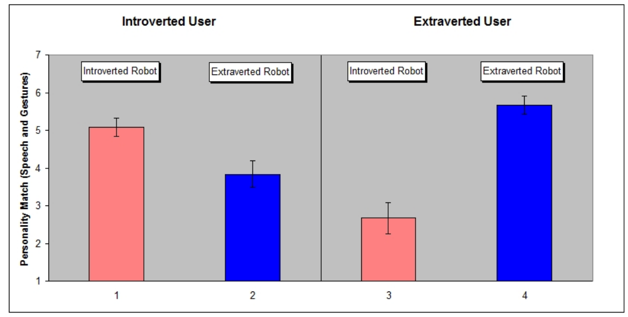


Fig. 6. Personality Matching for Introverted Robot and Extraverted Robot Conditions

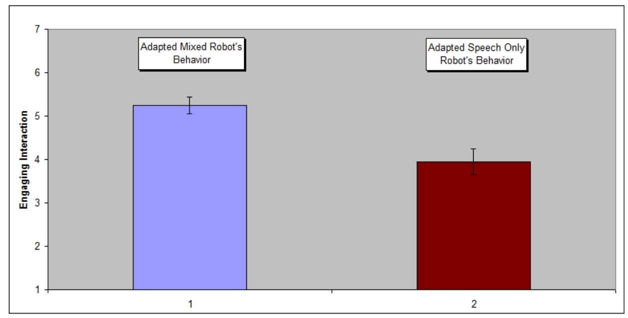


Fig. 7. Engaging Interaction: Adapted Mixed vs. Adapted Speech Only Robot Behavior Conditions

Moreover, all participants found significantly, but marginally more personality traits in the extraverted robot than in the introverted robot ($F[1,41] = 4.1583$, $p = 0.048$).

For the second hypothesis, the participants tested two other conditions: adapted combined robot behavior (gestures and speech adapted to the user's profile and personality) and adapted speech only robot behavior. Participants found the interaction with the adapted combined robot behavior more engaging than the adapted speech only robot behavior ($F[1,41] = 13.16$, $p = 0.0008$) (Figure 7). Participants were also asked if the robot behavior was appropriate. Through ANOVA, we found that the adapted speech only robot behavior was significantly considered less appropriate than the adapted combined robot behavior ($F[1,41] = 20.16$, $p = 0.0$). Furthermore, this result was also observed when the analysis was performed separately on the introverted participants ($F[1,23] = 9.5422$, $p = 0.005361$) and on the extraverted participants ($F[1,17] = 10.5625$, $p = 0.005$).

Our results showed that personality plays an important role in the interaction and that there is a difference of perception and preference for the robot as a function of its personality. The introverted individuals preferred interacting with the introverted robot more than with the extraverted robot. The same tendency of user-robot personality matching was also observed for the extraverted individuals. Moreover, we emphasized that a robot behavior that mixes gestures and speech is more engaging and considered more appropriate and natural by the participants than a robot behavior based only on speech.

The reported effect of the experiments under different personality conditions is related to the variation in the verbal content and gestural characteristics generated by the robot. For example in the introverted condition, the robot didn't give lot of details about the restaurants, and the frequency of performing gestures was low, beside the low energy of gestures (this is due to PERSONAGE generator which receives a previously set personality score as introverted unlike the second hypothesis's conditions, where the Personality Recognizer toolkit is employed online in order to calculate the score of human personality (i.e., extraversion trait), regardless of any prescribed personality scores as in the first hypothesis conditions, then generates a verbal content corresponding to that personality trait's score. In parallel, BEAT toolkit generates a set of gestures corresponding to the generated verbal content by PERSONAGE generator. The control parameters of the generated gestures, like: Speed, Amplitude, etc. were fixed experimentally through the range of the personality score from 10% (maximum introversion) to 100% (maximum extraversion). So that the robot implements the generated gestures by BEAT toolkit in a corresponding manner to the desired personality to show). The effect of this introverted behavior gave the extraverted participants, for example, the feeling that the reflected robot behavior doesn't tend a lot to be social or active like them when they speak or act, and that was the evaluated effect at the end through the questionnaires.

## VII. Conclusion

The paper describes a complete architecture for generating a combined verbal and nonverbal robot behavior based on the interacting human's personality traits. The personality dimensions of the interacting human are estimated through a psycholinguistic analysis of speech content. Furthermore,PERSONAGE generator uses the calculated personality scores in order to generate a corresponding text adapted to the interacting human's personality. Afterwards, BEAT toolkit is used in order to generate different kinds of gestures corresponding to the input text (in parallel with our developed general metaphoric gesture generator, which generates gestures based on the human's speech). Our work proves the important role of human-robot personality matching in creating a more appropriate interaction, and shows that the adapted combined robot behavior expressed through gestures and speech is more engaging and natural than the adapted speech-only robot behavior. Besides, this paper proves that extraverts prefer high speed robot movements contrarily to introverts, and that the perceived semantic matching between the generated robot speech and gestures, was higher in the extraverted condition than in the introverted condition. For the future work, we are interested in realizing a more dynamic synchronization between the affiliate and the stroke phase. Besides, we are interested in extending PERSONAGE language generator to include other domains than tourism and restaurants.

## Acknowledgment

This work was supported by the French National Research Agency (ANR) through Chaire D'Excellence program 2009 (Human Robot Interaction for Assistive Applications). The project website is: http://www.ensta-paristech.fr/~tapus/HRIAA/.

## References

[1] B. Grosz, "Team: A transportable natural language interface system," in Proceedings of the Conference on Applied Natural Language Processing, Santa Monica, CA, USA, 1983, pp. 39–45.

[2] T. Finin, A. Joshi, and B. Webber, "Natural language interactions with artificial experts," in Proceedings of the IEEE,10(2), 1986, pp. 921–938.

[3] W. Wahlster and A. Kobsa, "User models in dialog systems." Berlin,Germany: Springer Verlag, 1989, pp. 4–34.

[4] I. Zukerman and D. Litman, "Natural language processing and user modeling: Synergies and limitations," in User Modeling and User-Adapted Interaction, vol. 11, no. 1-2, 2001, pp. 129–158.

[5] E. Andre, T. Rist, S. Mulken, M. Klesen, and S. Baldes, "The automated design of believable dialogues for animated presentation teams," in S. Prevost J. Cassell, J. S. and Churchill, E., editors,Embodied conversational agents. Cambridge, MA, USA: MIT Press, 2000, pp. 220–255.

[6] J. Cassell and T. Bickmore, "Negotiated collusion: Modeling social language and its relationship effects in intelligent agents," in User Modeling and User-Adapted Interaction, vol. 13, 2003, pp. 89–132.

[7] K. Forbes-Riley and D. Litman, "Investigating human tutor responses to student uncertainty for adaptive system development," in Lecture Notes in Computer Science, vol. 4738, 2007, pp. 678–689.

[8] K. Forbes-Riley, D. Litman, and M. Rotaru, "Responding to student uncertainty during computer tutoring: An experimental evaluation," in Lecture Notes in Computer Science, vol. 5091, 2008, pp. 60–69.

[9] V. Ng-Thow-Hing, P.Luo, and S. Okita, "Synchronized gesture and speech production for humanoid robots," in Proc. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Taipei, Taiwan, Oct. 2010.

[10] S. Kopp, K. Bergmann, and I. Wachsmuth, "Multimodal communication from multimodal thinking - towards an integrated model of speech and gesture production," Semantic Computing, vol. 2, no. 1, pp. 115–136, 2008.

[11] E. Bevacqua, M. Mancini, and C. Pelachaud, "Speaking with emotions," in AISB Convention: Motion, Emotion and Cognition, University of Leeds, UK, 2004.

[12] B. Hartmann, M. Mancini, and C. Pelachaud, "Formational parameters and adaptive prototype instantiation for mpeg-4 compliant gesture synthesis," in Proceedings of Computer Animations, IEEE Computer Society Press, 2002.

[13] L. R.Goldberg, "An alternative description of personality: The big-five factor structure," Journal of Personality and Social Psychology, vol. 59, pp. 1216–1229, 1990.

[14] H. Eysenck, "Dimensions of personality: 16, 5 or 3? criteria for a taxonomic paradigm," Personality and Individual Differences, vol. 12, pp. 773–790, 1991.

[15] I. M. Briggs and P. Myers, "Gifts differing: Understanding personality type," Davies-Black Publishing, 1980.

[16] H. Christensen and E. Pacchierotti, "Embodied social interaction for robots," in Artificial Intelligence and Simulation of Behavior Convention (AISB'05), K. Dautenhahn, Ed., Hertsfordshire, UK, 2005, pp. 40–45.

[17] J. Eriksson, M. J. Matarić, and C. Winstein, "Hands-off assistive robotics for post-stroke arm rehabilitation," in Proc. IEEE International Conference on Rehabilitation Robotics (ICORR'05), Chicago, Il, USA, Jun. 2005, pp. 21–24.

[18] C. Nass and M. K. Lee, "Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency attraction," Journal of Experimental Psychology: Applied, vol. 7, pp. 171–181, 2001.

[19] S. N. Woods, K. Dautenhahn, C. Kaouri, R. te Boekhorst, K. L. Koay, and M. L. Walters, "Are robots like people? - relationships between participant and robot personality traits in human-robot interaction studies," Interaction Studies, vol. 8, no. 2, pp. 281–305, 2007.

[20] A. Tapus and M. J. Matarić, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance," in Proceedings of the AAAI Spring Symposium on Emotion,Personality and Social Behavior, Palo Alto, CA, USA, 2008.

[21] B. Reeves and C. Nass, "The media equation," University of Chicago Press, 1996.

[22] F. Mairesse, M. Walker, M. Mehl, and R. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," Journal of Artificial Intelligence Research (JAIR), pp. 457–500, 2007.

[23] F. Mairesse and M. Walker, "Controlling user perceptions of linguistic style: Trainable generation of personality traits," Journal of Computational Linguistics, vol. 37, 2011.

[24] J. Cassell, H. Vilhjalmsson, and T. Bickmore, "Beat: The behavior expression animation toolkit," in Proceedings of SIGGRAPH, 2001, pp. 477–486.

[25] A. Furnham, Language and Personality. In Giles, H. and Robinson, W. (Eds.), Handbook of Language and Social Psychology. Winley, 1990.

[26] J. Dewaele and A. Furnham, "Extraversion: the unloved variable in applied linguistic research," Language Learning, vol. 49, no. 3, pp. 509–544, 1999.

[27] K. Scherer, Language and Personality. In Scherer, K. R., and Giles, H. (Eds.), Social markers in speech, Cambridge University Press, 1979.

[28] J. Pennebaker and L. King, "Linguistic styles: Language use as an individual difference," Journal of Personality and Social Psychology, vol. 77, pp. 1296–1312, 1999.

[29] M. Mehl, S. Gosling, and J. Pennebaker, "Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life," Journal of Personality and Social Psychology, vol. 90, pp. 862–877, 2006.

[30] M. Coltheart, "The mrc psycholinguistic database," Quarterly Journal of Experimental Psychology, vol. 33, no. A, pp. 497–505, 1981.

[31] A. Gill and J. Oberlander, "Taking care of the linguistic features of extraversion," in Proceedings of the 24th Annual Conference of the Cognitive Science Society, 2002, pp. 363–368.

[32] K. Vogel and S. Vogel, "L'interlangue et la personnalite de l'apprenant," International Journal of Applied Linguistics, vol. 24, no. 1, pp. 48–68, 1986.

[33] E. Reiter and R. Dale, Building Natural Language Generation Systems. Cambridge University Press, 2000.

[34] A. Stent, R. Prasad, and M. Walker, "Trainable sentence planning for complex information presentation in spoken dialog systems," in Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), 2004, pp. 79–86.

[35] C. Fellbaum, WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA, 1998.

[36] T. Chklovski and P. Pantel, "Verbocean: Mining the web for fine-grained semantic verb relations," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2004.

[37] B. Lavoie and O. Rambow, "A fast and portable realizer for text generation systems," in Proceedings of the 3rd Conference on Applied Natural Language Processing, 1997, pp. 265–268.

[38] A. Aly and A.Tapus, "Speech to head gesture mapping in multimodal human-robot interaction," in Proceedings of the European Conference on Mobile Robotics (ECMR), Orebro, Sweden, 2011.

[39] ——, "An integrated model of speech to arm gestures mapping in human-robot interaction," in Proceedings of the 14th IFAC Symposium on Information Control Problems in Manufacturing (INCOM), Romania, 2012.

[40] A. Kendon, Gesticulation and speech: two aspects of the process of utterance, 1980.

[41] G. Matthews, I. Deary, and M. Whiteman, "Personality traits," Cambridge University Press, 2003.