HAL
open science

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

Amir Aly, Adriana Tapus

To cite this version:

Amir Aly, Adriana Tapus. A Model for Synthesizing a Combined Verbal and Nonverbal Behavior Based on Personality Traits in Human-Robot Interaction. The 8th ACM/IEEE Human-Robot Interaction Conference (HRI), Mar 2013, Tokyo, Japan. hal-01284708v1

HAL Id: hal-01284708

https://hal.science/hal-01284708v1

Submitted on 8 Mar 2016 (v1), last revised 6 Feb 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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*—In Human-Robot Interaction (HRI) scenarios, an intelligent robot should be able to synthesize an appropriate behavior adapted to human profile (i.e., personality). Recent research studies discussed the effect of personality traits on human verbal and nonverbal behaviors. The dynamic characteristics of the generated gestures and postures during the nonverbal communication can differ according to personality traits, which similarly can influence the verbal content of human speech. This research tries to map human verbal behavior to a corresponding verbal and nonverbal combined robot behavior based on the extraversion-introversion personality dimension. We explore the human-robot personality matching aspect and the similarity attraction principle, in addition to the different effects of the adapted combined robot behavior expressed through speech and gestures, and the adapted speech-only robot behavior, on interaction. Experiments with the NAO robot are reported.**

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. Grosz [38] tried to create a limited verbal natural language interface in order to access information in a database. An interesting theoretical study on the Natural Language (NL) was discussed in Finin et al. [30], in which they 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 Wahlster and Kobsa [95] and Zukerman and Litman [102], where they focused on the field of user modeling (i.e., understanding the user's beliefs, goals, and plans) in artificial intelligence dialog systems, and illustrated the importance of such modeling on interaction. Later on, some research studies discussed how persuasive will be the dialogue systems that are adapted to the user's model (including the ability to change explicitly and dynamically the aspects of the relationship with the interacting human through the use of social talks in the same way as humans behave) [7, 16, 31, 32].

Some efforts were driven towards generating synchronized verbal and nonverbal behaviors as discussed in Ng-Thow-Hing et al. [75]. The authors presented a system able to synchronize expressive body gestures with speech. This model was implemented on Honda humanoid robot (ASIMO), and was able to synthesize gestures of different types, such as: iconic, metaphoric, deictic, and beat gestures [64, 65]. Le and Pelachaud [53] discussed an interesting system for synthesizing co-speech and gestures for the NAO robot. They used the SAIBA framework [49] in order to generate a multimodal behavior designated to virtual agents, then they interfaced it with the NAO robot in order to generate and to model a synchronized verbal and nonverbal combined robot behavior. Similarly, virtual agents had received much attention concerning generating expressive behaviors. Kopp et al. [50] 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 manner. For example, it is sufficient to support the system with some a priori information about a certain object to describe, and the system will be able to generate itself an expressive verbal and nonverbal combined behavior exactly as humans do. Another interesting approach was discussed in Hartmann et al. [42], Bevacqua et al. [13], Mancini and Pelachaud [62], and Niewiadomski et al. [76]. The authors developed the virtual conversational agent GRETA, which uses verbal and nonverbal behaviors to express intentions and emotional states. It can be used as a dialog companion, a virtual tutor, a game-actor, or even a storyteller. Cassell et al. [17] introduced the conversational agent REA, which presents a real estate sales person through a multimodal expressive behavior. 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 a combined verbal and nonverbal

robot behavior based on the human's personality traits.

Personality is an important factor in human social interaction. In the related literature, there are different models of personality, such as: Big5 (*Openness, Conscientiousness, Extraversion-Introversion, Agreeableness, and Neuroticism*) [36, 37], Eysenck Model of Personality (PEN) (*P: Psychoticism, E: Extraversion, and N: Neuroticism*) [26, 27], and Meyers-Briggs (*Extraversion-Introversion, Sensation-Intuition, Thinking-Feeling, and Judging-Perceiving*) [70, 71]. In this research, the Personality Recognizer toolkit (Section III-A) integrated to our system is based on the Big5 personality model, as it is the most descriptive model of human personality. Morris [69], Dicaprio [23], Woods et al. [99], and Tapus and Matarić [90] defined personality as: "*the pattern of collective character, behavioral, temperamental, emotional and mental traits of an individual that has consistency over time and situations*". Consequently, it is obvious that the long term effect of personality on the generated behavior, makes it more reliable for characterizing the generated verbal and nonverbal behaviors, to the contrary of other short-term characteristics, like the prosodic features of speech.

Based on these findings, we assume that personality is an important factor within a human-robot interaction context. In this research, we try to develop a customized verbal and nonverbal combined robot behavior based on the extraversion-introversion personality trait of the interacting human. We focus on validating that the participants prefer interacting more with the robot when it has a similar personality to theirs, and that the adapted multimodal combined robot behavior (i.e., robot-user personalities match in terms of the type and level of the extraversion-introversion dimension, and that both speech and gestures are expressed synchronously) is more engaging than the adapted speech-only robot behavior (not accompanied with gestures). The context of interaction in this research is restaurant information request, in which the robot gives the required information about the selected restaurants to the interacting human in real-time, expressed through a combined verbal and nonverbal behavior [5].

The rest of the paper is structured as following: Section (II) discusses the importance of personality traits in human-robot interaction, Section (III) presents a general overview of the system architecture, Section (IV) describes the nonverbal behavior knowledge base extension, Section (V) illustrates how we realized the synchronized verbal and nonverbal behaviors on the robot, Section (VI) illustrates the hypotheses, design, and scenario of interaction, Section (VII) provides a description of the experimental results, Section (VIII) discusses the outcome of the study, and last but not least, Section (IX) concludes the paper.

II. WHY SHOULD PERSONALITY TRAITS BE CONSIDERED IN HUMAN-ROBOT INTERACTION?

In Human-Robot Interaction (HRI), a straightforward relationship has been found between personality and behavior [25, 74, 100]. 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, Reeves and Nass [81], Nass and Lee [74], and Tapus and Matarić [90] proved empirically that the human interacting with a dialog machine will spend more time on the assigned task if the system's behavior matches with his/her personality, which validates the similarity attraction principle (i.e., individuals are more attracted by others who have similar personality traits) in human-robot interaction situations [14]. Another interesting topic was discussed in Park et al. [77], in which they examined the influence of the KMC-EXPR robot personality (reflected only through facial expressions using the eyes and the mouth, with big movements for extraverts and small movements for introverts) on its anthropomorphism, friendliness, and social presence. The results showed that the participants assigned the extraverted robot a higher degree of anthropomorphism compared to the introverted robot. On the other hand for friendliness and social presence, the results shown that the extraverted participants considered the extraverted robot more friendly and more socially present than the introverted robot, while the introverted participants preferred more the introverted robot. These findings validate the similarity attraction principle [14].

Another interesting concept is the complementarity attraction (i.e., individuals are more attracted by others whose personalities are complementary to their own personalities) [45, 54, 89]. The effect of the AIBO robot personality on the interacting participants through relatively long-duration experiments, has been studied in Lee et al. [55]. The authors found that the participants preferred interacting more with the robot when it had a complementary personality than when it had a similar personality to their own personalities. Generally, the confusion between the similarity and complementarity attraction principles could be related to the context of interaction. Consequently, any of them could be validated during a human-robot interaction experiment, similarly to the human-human social attraction that involves either the similarity or the complementarity attraction during interaction [24]. For example, the similarity attraction looks more appropriate for the experimental design that considers the effect of the initial interaction between a human user and a robot on the developing relationship (which could be figured in most friendships between humans, where they get attracted to each other based on the matching between their personalities and the equality of dominance between each other). Meanwhile, the complementarity attraction contends more for long-term relationships (e.g., marriage and some kinds of friendship of different roles, where one person is more dominant than the other) [93]. In this research, we are interested in making the interacting human more attracted to the robot during the conducted experiments, so that the robot takes a similar personality to the interacting human's personality (i.e., similarity attraction principle is being examined). Furthermore, due to the relatively short-duration of the conducted experiments, the validation

of the complementarity attraction principle (using the current experimental design) would be hard to be accomplished.

A strong psychological evidence that firmly supports our focus on the similarity attraction principle, is the *chameleon effect*. This effect refers to the "*nonconscious mimicry of the postures, mannerisms, facial expressions, and verbal and nonverbal behaviors of one's interaction partners, such that one's behavior passively and unintentionally changes to match that of others in one's current social environment*", which happens frequently and naturally between people [19]. This definition matches the findings of Bargh et al. [9], which suggested that the perception of one's behavior enhances the chances of engaging in that behavior by his/her counterpart. Giles and Powesland [34] discussed mimicry in speech and found that people tend to mimic the accents of their interaction partners. Other speech characteristics like speech rate and rhythm are also mimicked during interaction [15, 96]. Similarly, Lafrance [51] and Bernieri [12] found that gestures, postures, and mannerisms are mimicked during interaction. This verbal and nonverbal behavior mimicry reported a higher positive effect on interaction than the cases when mimicry was absent [19]. Maurer and Tindall [63] found that the mimicry of a client's arm and leg positions by a counselor, increased the client's perception of the empathy level of the counselor. Van-Baaren et al. [92] found that when a waitress mimicked verbally her customers, she received a larger amount of tips. Bailenson and Yee [8] found that mimicking the participant's head movements by a virtual agent was perceived more convincing and was attributed a higher trait ratings than the non-mimicking interaction cases.

Moreover, several studies investigated the relationship between behavior mimicry and attraction. Gump and Kulik [40] discussed that behavior mimicry enhances the coherence within interaction by making the interacting partners look similar to each other. Gueguen [39] studied the effect of the verbal and nonverbal behavior mimicry on a courtship relationship. He found that the male participants preferred the female participants who mimicked them. Luo et al. [59] found that people preferred similar gestures to their own during human-agent interaction, which matches the outcome of the previous studies. Additionally and most importantly, this study suggested a *preliminary* relationship between personality and the perception of an exercised behavior. This last primary result in addition to all the previous discussion open the door to a more elaborate study that investigates the link between personality and behavior, which constituted a strong inspiration for our current research study.

On the other hand, Barrick and Mount [10] investigated the general relationship between personality and professions. They found that some professions, such as: teacher, accountant, and doctor, tend to be more introverted, while other professions, such as: salesperson and manager, tend to be more extraverted. A similar tendency was discussed in Windhouwer [98], which tried to investigate how could the NAO robot be perceived intelligent in terms of its profession and personality. They found that when the robot played the role of an introverted manager, it appeared more intelligent than the extraverted manager. Similarly, when the robot played the role of an extraverted teacher, it appeared more intelligent than the introverted teacher. These last findings oppose - to some extent - the findings of Barrick and Mount [10], which could be due to some differences in the context of interaction. For example, when the robot was playing the role of an introverted manager during a meeting, it probably seemed deeply thinking about work problems trying to reach optimal solutions. This could have given the introverted robot a more intelligent look than the extraverted robot that was not looking thinking enough and was moving fast with high energy. Therefore, the findings of Barrick and Mount [10] could be considered as general findings that could differ experimentally according to the context of interaction, which makes the matching between robot personality and profession (task), a difficult point to estimate in advance before experiments. However, it is worthy with study, as it can influence positively the way people perceive the robot.

Moreover, Leuwerink [56] discussed how people would perceive the robot intelligent in terms of its personality within dyadic and group interactions. They found that the introverted robot was perceived more intelligent in a group interaction. Meanwhile, the extraverted robot was perceived more intelligent in a dyadic interaction. These findings match the findings of Barrick and Mount [10] in a general manner for certain professions, such as: teacher for the introverted robot in a group interaction, and salesperson for the extraverted robot in a dyadic interaction. However as mentioned earlier, it all depends on the context of interaction, because an extraverted robot-teacher could be more suitable for a group interaction, considering that it will appear more active and funny. Therefore, it is difficult to draw a common definition for the relationship between personality, profession, and group/dyadic interaction due to the differences that may appear in each experimental study.

On the other hand, other studies found a relationship between human personality and proxemics (i.e., the study of the interpersonal distance's influence on interaction) [41, 91], which influences the robot navigation planners in human-robot interaction situations (e.g., extraverted people are more tolerant of their personal space invasion by a robot than introverted people) [97]. Nakajima et al. [72, 73] discussed the influence of emotions and personality on the social behaviors of human-robot collaborative learning systems. They found that the users had more positive impression about the usefulness of the learning experience when the cooperative agent displayed some social responses with personality and emotions. Generally, all the previous discussion reveals the feasibility of considering personality traits in human-robot interaction scenarios, which can attract humans to interact more efficiently with robots.

Several studies discussed the importance of the extraversion-

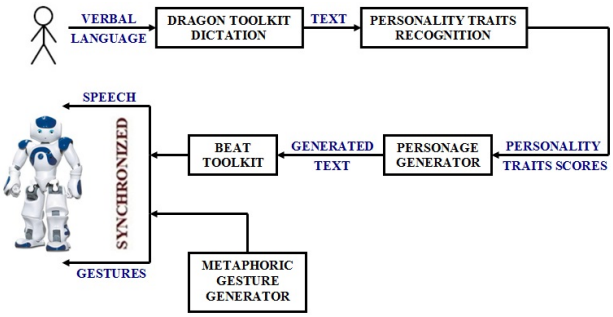

Fig. 1: General overview of the system architecture

introversion dimension in characterizing human behavior. J-Campbell et al. [46] and Selfhout et al. [87] discussed the important effect of both the agreeableness and the extraversion-introversion dimensions on developing human peer relationships. Lippa and Dietz [57] indicated that the extraversion-introversion dimension is the most influential and accurate trait among the Big5 personality dimensions. Besides, Moon and Nass [68], Isbister and Nass [45], and Nass and Lee [74] discussed the importance of the extraversion-introversion dimension in Human-Computer Interaction (HCI). On the other hand, several research studies considered the verbal and nonverbal cues as the most relevant cues for personality traits analysis [43, 74, 79, 85]. Consequently, this work tries to demonstrate the influence and the importance of personality in human-robot interaction contexts. It links between the extraversion-introversion dimension and the verbal and nonverbal behavioral cues, for the purpose of generating an adapted robot behavior to human personality so as to reinforce the level of interaction between a human user and a robot.

III. SYSTEM ARCHITECTURE

Our system is a coordination between different sub-systems: (1) Dragon Naturally Speaking toolkit, which translates the spoken language of the interacting human into a text, (2) Personality Recognizer, which estimates the interacting human's personality traits through a psycholinguistic analysis of the input text [61], (3) PERSONAGE natural language generator, which adapts the generated text to the interacting human's personality dimensions [60], (4) BEAT toolkit, which translates the generated text into gestures (not including the general metaphoric gestures) [18], (5) Metaphoric general gesture generator (Section V) [6], and (6) The humanoid NAO robot as the test-bed platform. An overview of the system architecture is illustrated in Figure (1).

A. Personality Recognizer

Personality markers in language had received a lot of interest from psycholinguistic studies. Scherer [86], Furnham [33], and Dewaele and Furnham [22] described how could the extraversion-introversion personality trait influence linguistically speech production. They stated that extraverts are more loud-voiced and talk more iteratively with less faltering and pauses, than introverts. Moreover, extraverts have high verbal output and speech rate, and use informal language, while introverts use a rich vocabulary. On the other hand, extraverts express more encouragement and agreement, and use more positive feeling words, than introverts [78].

A general approach for characterizing the majority of personality traits was discussed in Pennebaker and King [78], in which they used the Linguistic Inquiry and the Word Count toolkit (LIWC) in order to define the word categories of 2479 essays (containing 1.9 million words) written by different persons covering the five personality traits described in the Big5 Framework [36, 37]. This dictionary enabled them to state general relationships and characteristics for the five personality traits. Conscientious people -for example- avoid negative feeling words, negations, and words expressing discrepancies. Similarly, Mehl et al. [67] created a spoken data corpus (containing 97468 words and 15269 utterances) in addition to their transcripts, covering different personality traits. This corpus was sub-divided into several word categories using the LIWC tool.

The findings of the previous data corpora were the basic body of the research conducted by Mairesse et al. [61]. They created a huge database including the LIWC psycholinguistic features, such as: anger words (e.g., hate), metaphysical issues (e.g., god), and family members (e.g., mom, brother), in addition to other psycholinguistic features included in the MRC database [21], such as: frequency of use (e.g., low: nudity, duly and high: the, he) and concreteness (e.g., low: patience, high: ship), besides the utterance type features, such as: command (e.g., must, have to), prompt (e.g., yeah, ok), and question-assertion (which is any utterance out of the previous categories). The relationship between the utterance type features and personality traits was discussed in Vogel and Vogel [94] and Gill and Oberlander [35], 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 express several personality dimensions through language. The architecture of PERSONAGE generator is illustrated in Figure (2), which is based on the traditional pipelined natural language generation (NLG) architecture [82]. The input consists of personality traits' scores, besides the selected restaurant(s) in New York City. The database of PERSONAGE generator contains scalar values representing the ratings of 6 attributes (used for recommendation and/or comparison according to the experimental context): cuisine, food quality, service, atmosphere, price, and location of more than 700 restaurant collected from real surveys investigating the opinion of people visited these restaurants. The content of the generated language

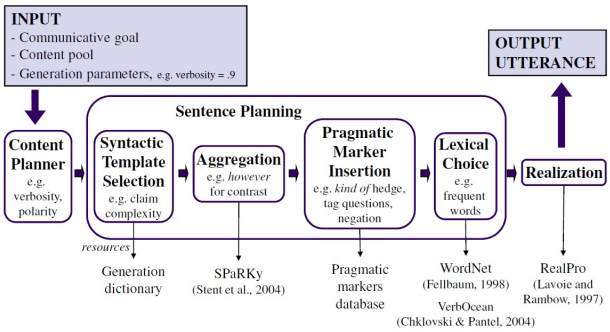

Fig. 2: Architecture of PERSONAGE generator [60]

could be more controlled through some parameters, like the verbosity parameter, which could be set to 1 in order to maximize the wordy content of the generated utterance.

The content planner plays the role of choosing and structuring (in a tree format) the necessary information to be processed by the sentence planner, in terms of the values of some parameters, such as: verbosity, polarity, and repetition (i.e., the content planer decides what to say). Meanwhile, the sentence planner deals with phrasing the information structured by the content planner. It searches in the dictionary, the group of primary linguistic structures attributed to 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, such as: verb, noun, etc.). Afterwards, it aggregates the obtained syntactic templates in order to generate a complete syntactic structure for the utterance [88].

On the other hand, the pragmatic marker insertion process in the sentence planner modifies the aggregated syntactic structure in order to generate several pragmatic effects, like: the hedge *you know*, the question tags, etc. The lexical choice process chooses the most appropriate lexeme (from many different lexemes expressed by PERSONAGE generator) for each word in terms of the frequency of use, and the lexeme's length and strength [20, 29]. Last but not least, the realization process which follows the sentence planner, transforms the resulting syntactic structure to a string using appropriate rules (e.g., the word insertion and morphological inflection rules) [52].

C. BEAT Toolkit

BEAT is the Behavior Expression Animation Toolkit that takes as an input a text and generates a corresponding synchronized set of gestures. It processes the contextual and linguistic information of the text so as to control body and face gestures, besides voice intonation. This mapping (from text to gesture) is implemented through a set of rules derived from intensive research on the nonverbal conversational behavior [18]. BEAT pipeline is composed of different XML-based modules, as illustrated in Figure (3). The language tagging module receives

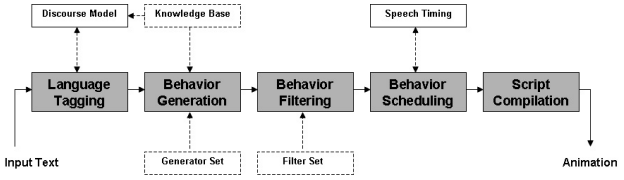

Fig. 3: Architecture of BEAT toolkit [18]

an XML tagged text generated from PERSONAGE generator, 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 gesture conflict and priority filters. The user-definable data structures, like: 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 important contextual information and definitions for generating relevant and precise nonverbal 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.
- **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.

IV. EXTENSION OF THE NONVERBAL BEHAVIOR KNOWLEDGE BASE OF BEAT TOOLKIT

The purpose of the performed extension on BEAT toolkit was to add necessary information about the generated text by PERSONAGE generator 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 restaurant's location, price category, size, and cuisine, which has been used in the interaction scenarios. Some instances were also added to the knowledge base describing some related places to the object "Restaurant", such as: "Basement" and "Dining Room"

in terms of their size, lightening, and painting. The new added scenes to the knowledge base define the restaurants' names, including the previously defined instances. The precised gestures' characteristics in the knowledge base concern different types of iconic gestures, including hand shapes and arm trajectories (unlike other gesture categories that do not require specific hand/arm shapes, as indicated in Section V). Some new linguistic keywords were aligned to specific iconic gestures with the corresponding hand/arm geometrical shapes' characteristics, like the adjective "narrow", which 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 semantically matches the adjective "narrow".

V. MODELING THE SYNCHRONIZED VERBAL AND NONVERBAL BEHAVIORS ON THE ROBOT

BEAT toolkit was built as a customizable gesture generator, so that more gesture categories could be added to the generation system of the toolkit, or even some extension could be imposed on its nonverbal behavior knowledge base in order to increase the expressivity scope of some built-in gestures (e.g., iconic gestures), as indicated in Section (IV). Generally, we found that the built-in gesture categories are mostly sufficient for the relatively short verbal context generated by PERSONAGE generator (except for the general metaphoric gestures, which are not included in BEAT toolkit. Therefore, we have integrated them externally to the system, as illustrated in Figure 5). In this research, we are interested only in four categories 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 specific allocated gestures. 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 gestures. The animation script reveals also that the adjective word "narrow" was attributed to an iconic gesture, where the two hands are used to model the gesture "gesture-both" (i.e., performing a gesture using both hands) with the shape "hands-in-front", 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 iconic gestures. They could take the form of a general hand/arm/head shaking or even a specific shape, like when we want to express a time sequence, we use the word "after" associated with a specific hand/arm motion symbolizing this idea. Therefore, this word

```
<?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: XML animation script for the generated verbal and nonverbal combined behavior

(in addition to other similar new words) was added and allocated in the knowledge base to the corresponding specific hand/arm motion trajectory, similarly to iconic gestures. On the other hand, the generation of general metaphoric gestures does not 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 so as to integrate the paraverbal modality into the generation of a nonverbal behavior, as will be explained in details later on.

The mapping of gaze, posture-shift, iconic, and specific-shape-metaphoric gestures from the animation script (Figure 4) to the robot, necessitates that the robot processes each line of the script indicating the duration of each chunk that contains a synchronized verbal content with an attributed nonverbal behavior. Kendon [48] defined *gesture phrases* as the primary units of gestural movement that include consecutive 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 the robot (in case of iconic and specific-shape-metaphoric gestures), is the required high temporal synchronization between the stroke (i.e., the expressive gesture phase) and the affiliate (i.e., the affiliated word or sub-phrase) in order to express an idea accurately. The time estimation indicated in the animation script reveals the calculated time for the stroke phase of 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 synchronously with the affiliate. Therefore, the gesture's stroke phase is fixed to lead the affiliate's onset by an approximate duration of one syllable (i.e., 0.3s).

Figure (5) illustrates the gestural behavior control architecture. The general metaphoric gesture generator receives as an input, the temporally aligned text with speech using a TTS (text-to-speech) engine, so that it synthesizes general metaphoric gestures corresponding to each word of the text based on the prosodic cues of the aligned speech segments to words [2, 3, 4, 6]. Consequently, the final chunks in the behavior controller would contain both the temporal and the corresponding word-index information of five gesture types (i.e., general metaphoric gestures, gaze gestures, posture-shift gestures, iconic gestures, and specific-shape-metaphoric gestures).

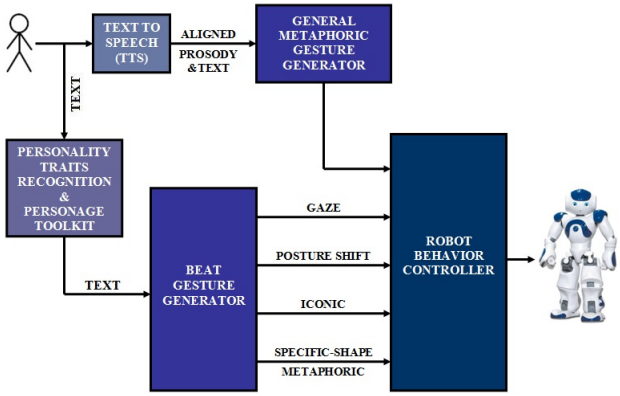

Fig. 5: Behavior control architecture

General metaphoric gestures are synthesized using the Coupled Hidden Markov Models (CHMM), which could be considered as a multi-stream collection of parallel HMM characterizing the segmented data of both prosody and gestures. The generated gestures are characterized by the most likely path of observations through the gesture channel of the CHMM (which is modeled in terms of the linear velocity and acceleration observations of body segments, in addition to the linear position observations of body articulations), given an observed audio sequence [6] (Appendix A). The inverse kinematics is applied on the generated linear position coordinates in order to calculate the corresponding rotation values of body articulations. Using the CHMM in generating metaphoric gestures allows for synthesizing gestures of varying amplitude and duration adapted to the human's prosodic cues. Besides, the random variations of the synthesized gestures' motion patterns make them look as natural as human gestures, which will not be the case if a fixed-gestures dictionary is employed instead. This methodology clarifies the quantitative difference between the generated amount of gestures in case of the introverted and extraverted conditions. Therefore, for an introverted speaker who does not speak a lot, he/she will have a corresponding limited pitch-intensity contours, which will lead to a corresponding limited set of generated gestures, contrarily to the extraverted individuals.

On the other hand, in order to reasonably reflect a specific introverted or extraverted personality on the robot, the generated motion curves' values of the synthesized gestures should be controlled in both personality conditions. Consequently, we attributed experimentally 10% of the amplitude of the generated motion curves' values to the maximum introversion level, while we kept 100% of the amplitude for the maximum extraversion level (based on the fact that the training database of the CHMM is depending on highly extraverted actors) [6]. The corresponding motion curves' values to the range of personality scores between the maximum introversion and extraversion levels (i.e., between 10% and 100%) could be easily derived as a function of the motion curves' values calculated at the maximum introversion and extraversion levels.

Unlike the automatic modeling of the synthesized general metaphoric gestures on the robot directly, the modeling of the other four types of gestures generated by BEAT toolkit was controlled inside the robot behavior controller. During the gaze gesture (whether it is oriented *towards the hearer* or *away from the hearer*), the whole neck turns so as to get oriented away/towards the interacting human (Figure 4). The neck movement was previously programmed (same for the posture-shift gesture in the directions: *lean forward* and *lean backward*). A similar tendency was applied for the generated iconic and specific-shape-metaphoric gestures, in which corresponding body movements to certain words in the knowledge base were also previously programmed. The control motion parameters of the generated gestures by BEAT toolkit have initially been set experimentally through the normal range of personality scores, from 10% (maximum introversion) to 100% (maximum extraversion) with a step of 10%, so that the robot implements the generated gestures in a corresponding approximate manner to the desired personality type and level to show. However, the encountered difficulty was to keep the temporal alignment between the generated gestures and text indicated in the animation script in Figure (4). Therefore, the robot behavior controller should be updating the time-control parameter of the programmed gestures based on their estimated duration in the animation script so as to make the robot finishes performing a specific gesture at the specified time instants in the script.

After designing the nonverbal behaviors corresponding to the five gesture types explained earlier, the robot behavior controller examines any existing conflict between the synthesized gestures. If there exists a conflict between an iconic or a specific-shape-metaphoric gesture (less frequent) and a general-hand/arm-metaphoric gesture (more frequent), so that both have to be implemented at the same time, the priority would be given automatically to the iconic or the specific-shape-metaphoric gesture. A similar tendency happens if a conflict occurs between a gaze gesture (in the direction *away from the hearer*) and a general-head-metaphoric gesture, in which the priority goes to the gaze gesture.

VI. EXPERIMENTAL SETUP

In this section, we introduce first the robot used in the experiments, then we follow by an overview for the conducted experiments.

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 eight full-color RGB eye leds, two cameras, an inertial sensor, a sonar sensor, and many other sensors that allow for perceiving the surrounding environment with high precision and stability.

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

B. Hypotheses

The presented research aim to test and validate the following hypotheses:

- **H1:** The robot behavior that matches the user's personality expressed through combined speech and gestures will be preferred by the user.
- **H2:** 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 two robot personalities:

- The robot uses introverted cues expressed through combined gestures and speech in order to communicate with the user.
- The robot uses extraverted cues expressed through combined gestures and speech in order to communicate with the user.

Similarly, in order to validate the second hypothesis, the user tested two different conditions:

- The robot communicates with the user through combined gestures and speech (the robot-user personalities match in terms of the type and the level of personality). We call it: adapted combined robot behavior.
- The robot communicates with the user only through speech (the robot-user personalities match in terms of the type and the level of personality). We call it: adapted speech-only robot behavior.

All the previous four conditions were randomly ordered during the experimental phases. For the second hypothesis, we excluded the condition of interaction through gestures only, as it does not fit in the normal context of the non-mute human-human interaction. Similarly, we excluded the condition of interaction through adapted speech and non-adapted gestures to human personality, because of the following reasons: (1) The production of human gestures and speech follows the same process, so that they are naturally aligned, (2) The characteristics of the naturally aligned human speech and gestures are both adapted to his/her personality; therefore, the generated robot speech and gestures should be both adapted to the interacting human's personality so as to make the interaction more engaging. Consequently, it is neither normal nor natural to consider that speech could be adapted to human personality alone without gestures (similarly to the adapted gestures and non-adapted speech interaction condition, which has not been considered in our study).

Generally, the main objective of the second hypothesis is to evaluate the importance of using adapted combined speech and gestures together during communication (instead of using adapted speech only) in order to better express and reflect ideas. In our experiments, we focused only on the extraversion-introversion dimension that indicates the level of sociability of an individual. An extraverted individual tends to be sociable, friendly, fun loving, active, and talkative, while an introverted individual tends to be reserved, inhibited, and quiet (Figures 6 and 7).

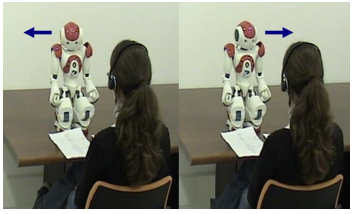

Fig. 6: Introverted robot condition (the robot gaze was more down-directed with a low gesture rate. The arrows refer to the direction of the head movement)

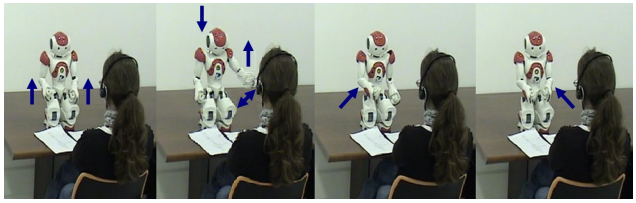

Fig. 7: Extraverted robot condition (the robot head was looking up-down, and the general metaphoric gesture rate of the head and the arms was high. The arrows refer to the direction of the head and arm movements)

The theme of the interaction in our experiments is restaurant information request. The robot has a list of restaurants in New York City, and its role is to give appropriate information about six elements: cuisine, food quality, service, location, atmosphere, and price for the selected restaurants in comparison. Our interaction scenario is described as following:

- The robot introduces itself as a guide to the participant and asks him/her to say somethings he/she knows about New York City. This first step is necessary for the robot in order to be capable of automatically identifying the participant's personality based on the analyzed linguistic cues.
- The robot has a list of restaurants and asks the participant to choose some restaurants so as to find out more details about them.
- The robot waits for the participant's input so as to produce appropriate combined speech and gestures based on the calculated personality traits.
- The participant asks for information about two restaurants of his/her choice.
- The robot gives the required information through a combined verbal and nonverbal behavior to the participant in real time.
- The participant can ask for more details about other

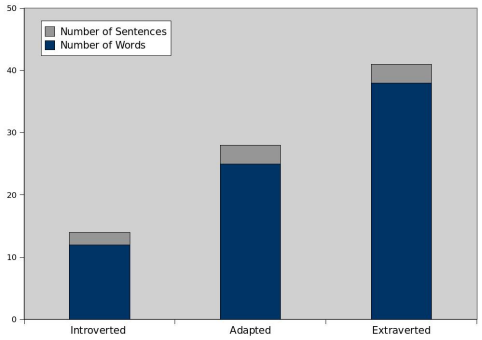

Fig. 8: Statistics of the synthesized words and sentences in Example 1, expressed in different personality conditions (the number of sentences is represented only by the gray areas, while the number of words is represented only by the blue areas)

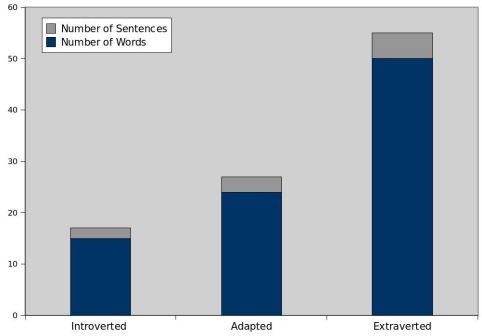

Fig. 9: Statistics of the synthesized words and sentences in Example 2, expressed in different personality conditions (the number of sentences is represented only by the gray areas, while the number of words is represented only by the blue areas)

restaurants (if he/she wants), and the robot gives back the required information, correspondingly.
- The interaction ends when the participant does not want to know more information about other restaurants, so that he/she has got the required information, they were searching for.

The following examples indicate the differences between the generated verbal output of PERSONAGE generator during the experimental phases, in which the robot gives information to the human about the compared restaurants in question:

Example 1: The statistics of the generated words and sentences in this example are summarized in Figure (8).

- **Introverted Personality**: *America is rather excellent. However, Alouette does not provide quite good atmosphere.*
- **Extraverted Personality**: *Alouette is an expensive bistro (French place in Manhattan), and it offers bad atmosphere and bad stuff. Alva provides nice service, the atmosphere is poor though. It is a new American place located near Union Square, you know.*
- **Adapted Personality**: *Amarone offers acceptable food, however, the atmosphere is poor. It has friendly waiters, but it is expensive. Although Alva is costly, the food is adequate.*

Example 2: Similarly to the previous example, the statistics of the synthesized words and sentences are summarized in Figure (9).

- **Introverted Personality**: *Acappella has nice food with quite outstanding waiters. However, Acacia does not have friendly waiters.*
- **Extraverted Personality**: *Even if Pho Bang has bad waiters and bad atmosphere, the food is nice. It is a small Vietnamese place in Manhattan. Even if Willow is expensive, the atmosphere is nice, you know. It is a new American place in Milltown. Also, this place offers nice service and nice food.*
- **Adapted Personality**: *Above has adequate waiters, also it offers decent food and pleasant atmosphere. Acacia provides acceptable food and friendly waiters. Its price is 40 USD.*

Example 3: Finally, the statistics of the synthesized words and sentences are summarized in Figure (10).

- **Introverted Personality**: *Alfama provides quite good atmosphere and rather outstanding stuff. While, Bar Odeon does not have nasty food.*
- **Extraverted Personality**: *America has bad stuff and bad atmosphere. Its price is* 27 *USD and it offers poor food. It is a new American place located near Union Square. Bar Odeon provides nice food, even if it is not expensive. Even if this place has rude waiters, the atmosphere is nice, you know. It is a French place in Manhattan.*
- **Adapted Personality**: *Jing Fong's price is* 21 *USD. This place which offers adequate food, is a big Chinese place. Bar Odeon has a pleasant atmosphere. Even if its price is* 44 *USD, the food is acceptable.*

The previous examples reveal the verbal content change of the generated utterances during the experimental phases. The formulation of the generated sentences could be manipulated through the tuning parameters of PERSONAGE generator. This variation made the participants feel that the robot is expressing more details in the extraverted condition than in the introverted condition, also it clarified the verbal content difference between the adapted personality condition from one side, and the other personality conditions (i.e., introversion and extraversion) from the other side.

The average duration of a single interaction in a given condition was varying between around 3 and 4 minutes. The system was evaluated based on user introspection (i.e., questionnaires). At the end of each experimental phase, each

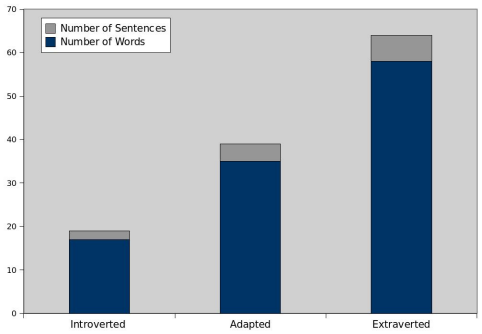

Fig. 10: Statistics of the synthesized words and sentences in Example 3, expressed in different personality conditions (the number of sentences is represented only by the gray areas, while the number of words is represented only by the blue areas)

participant completed one questionnaire designed to evaluate and judge: the synchronization between the generated robot speech and gestures, the human's impression about the reflected robot personality, the interaction with the robot, etc. All questions (i.e., 24 question) were presented on a 7-point Likert scale.

VII. EXPERIMENTAL RESULTS

The subject pool consisted of 21 participant (14 male, 7 female; 12 introverted and 9 extraverted). Introversion and extraversion are considered belonging to the same personality continuum scale; consequently, having a high score in one of them means having a corresponding complementary low score in the other one. Young [101] and Jung et al. [47] proposed a middle group of people in-between introverts and extraverts, called ambiverts, who have both introverted and extraverted features. The ambiversion range on the extraversion-introversion personality scale is equally distributed over the extraversion-ambiversion and ambiversion-introversion intervals. Supposing an ideal ambivert score is equal to 50%; therefore, we considered the participants with at least 25% introverted functions (i.e., with score less than or equal to 37.5%) to be introverted. Similarly, we considered the participants with at least 25% extraverted functions (i.e., with score greater than or equal to 62.5%) to be extraverted. In this study, all of the calculated personality scores were not included in the considered ambiversion interval (i.e., between 37.5% and 62.5%). Therefore, our analysis focuses only on two categories of participants: introverts and extraverts.

The experimental design was based on the within-subjects design, which has a probable carryover effect as a weak point, unlike the between-subjects design. However, the reasons for choosing it for the experimental setup are: (1) To minimize the variance error associated with the individual differences of the participants, so that the participants were the same in each experimental phase, (2) It is so difficult to recruit four times the actual number of participants (i.e., 84 participant) in order to validate the between-subjects design of the conducted experiments (two experimental hypotheses, where each one contains two phases). The four experimental phases validating the stated experimental conditions in Section (VI-C), were randomly ordered. The recruited participants were ENSTA-ParisTech undergraduate and graduate students, whose ages varied between 21-30 years old.

In order to test the first hypothesis, all the participants were exposed to two conditions: introverted robot and extraverted robot. In the introverted robot condition, the generated robot gestures were narrow, slow, and executed at a low rate. Contrarily, in the extraverted condition, the generated robot gestures were broad, quick, and executed at a high rate (Section V). The generated speech content is also based on personality; the robot gave more details in the extraverted condition than in the introverted condition.

Our ANOVA analysis showed that the extraverted individuals perceived the extraverted robot as significantly more close to their personality than the introverted robot ($F[1,17] = 40.5$, $p < 0.01$). 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 11). All the participants (introverted and extraverted together) considered that the robot speech and gestures were semantically matched (i.e., there was a matching in the meaning of both speech and gesture content based on the participants' observations) [11, 64, 65, 66], significantly more in the extraverted condition than in the introverted condition ($F[1,41] = 9.29$, $p = 0.0041$). However, when the user's extraversion-introversion personality trait was included in the analysis, this aspect was significant only for the extraverted individuals ($F[1,17] = 6.87$, $p = 0.0185$).

When the participants were asked about their preference for the speed of gestures, the extraverted users preferred the extraverted robot with fast movements to the introverted robot ($F[1,17] = 9.71$, $p = 0.0066$) (Figure 12), while the introverted users preferred the introverted robot with slow movements to the extraverted robot ($F[1,23] = 16.65$, $p = 0.0005$). These findings are in concordance with the findings of Eysenck [26, 27] and Eysenck and Eysenck [28], which linked the extraversion-introversion personality dimension to the activity level, considering the high activity level as an extraverted feature, meanwhile the low activity level tends more to characterize introversion.

For the second hypothesis, two other conditions have been examined with all the participants: adapted combined robot behavior (i.e., gestures and speech are adapted to the user's extraversion-introversion personality trait), and adapted speech-only robot behavior. The participants found the adapted combined robot behavior more engaging than the adapted speech-only robot behavior ($F[1,41] = 13.16$, $p = 0.0008$) (Figure 13). Through ANOVA test, we found that the adapted speech-only robot behavior was significantly considered less appropriate ($F[1,41] = 20.16$, $p < 0.01$) and less social

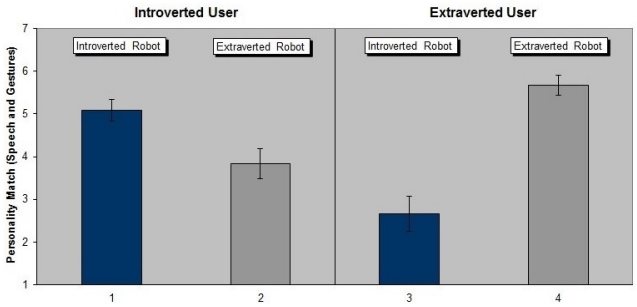

Fig. 11: Personality matching for the introverted and extraverted robot conditions

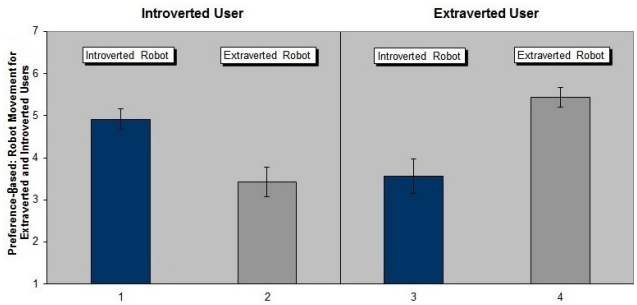

Fig. 12: Preference of the introverted and extraverted users for the robot movement

($F[1,41] = 9.137$, $p = 0.004$) than the adapted combined robot behavior. Moreover, the participants (i.e., the introverted and extraverted participants together) found that the execution of arm movements was fluid with an average score of $M = 4.2$ on a 7-point Likert scale (fluidity is an independent feature of the extraversion-introversion effect on gesture characteristics). At the same time, they agreed that the robot speech and gestures were semantically matching, and that they were well synchronized with average scores of $M = 5.05$, $SD = 0.59$ and $M = 4.96$, $SD = 0.74$, respectively. The participants agreed that the combined use of speech and gestures appeared natural with an average score of $M = 4.72$, $SD = 0.95$. On the other hand, when asked if the robot was helpful, no significant difference was found between the adapted speech-only and the adapted combined robot behaviors, with average scores of $M = 5.19$, $SD = 1.36$ and $M = 5.57$, $SD = 1.54$, respectively. The previous results confirm that personality plays an important role in interaction, so that it controls both the human's perception and preference for the robot, which makes it an important factor to consider in human-robot interaction contexts.

VIII. Discussion

In this study, we investigated the similarity attraction principle within a human-robot interaction scenario, in which the robot adapts its multimodal combined behavior to the interacting human's personality, and we explored the perception of the interacting human for the generated behavior. Moreover, we investigated the effect of the multimodal combined robot behavior expressed through speech and gestures on interaction, compared to the single-modal robot behavior expressed only through speech.

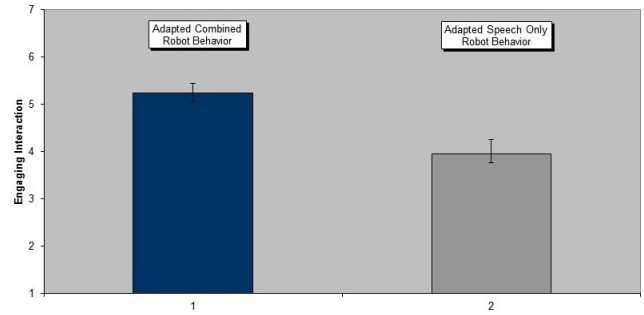

Fig. 13: Engaging interaction: adapted combined and adapted speech-only robot behavior conditions

The obtained results validated that the behavior of the robot was more preferred when it got adapted to the interacting human's personality. Figure (11) illustrates the human's personality-based preference for the robot behavior, and reveals the binary perception of the extraverted users for the introverted and extraverted robot conditions. To the contrary, some of the introverted users had a remarkable preference for the extraverted condition of the robot, however this preference was not dominant, so that the similarity attraction principle was validated. This variance in the perception of the robot behavior between the introverted and extraverted participants, reveals the difficulty in setting up clear borders that could separate experimentally the similarity and complementarity attraction principles. We argue that both of the similarity and complementarity attractions could co-exist during interaction, so that any of them could be validated based on the context and conditions of interaction. However, this needs an elaborate study and a large number of participants for validation. These last findings are consistent with the findings of the previous studies in the related literature discussed in Section (II), which validated *basically* the similarity attraction principle and stated some adverse cases, where the complementarity attraction principle was valid. This matches our proposed argument that any of the two attraction principles could be validated according to the context and conditions of interaction.

On the other hand, the results proved the important role of the multimodal robot behavior in making the interaction more engaging than the interaction that involves single-modal robot behavior (Figure 13). This logical result opens the door to other broader studies that employ more communicative cues like facial expressions so as to investigate and compare between the effects of different single and combined modalities of communication on interaction. The previous findings stating the positive effect of the robot behavior multimodality on interaction, are consistent with the findings of most of the related state-of-the-art studies [44, 58].

IX. 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 in order to include other domains than tourism and restaurants.

Appendix

A. General Metaphoric Gesture Generation

Our proposed system for synthesizing metaphoric gestures is integrated through 3 stages, as illustrated in Figure (A1) [1, 6]. Stage 1 constitutes the training phase of the system, through which the raw speech and gesture training inputs get processed in order to extract relevant features (e.g., the pitch-intensity curves for speech and the motion curves for gesture). Afterwards, the calculated characteristic curves undergo both of the segmentation phase (which is concerned with segmenting a continuous sequence of gestures into independent gestures using the kinetic features of body segments, and with segmenting speech into corresponding syllables to the segmented gestures, for which their prosodic cues will be calculated), and the Coupled Hidden Markov Models (CHMM) training phase. The segmented patterns of prosody and gestures are modeled separately into two parallel HMM constituting the CHMM [80, 83, 84], through which new metaphoric head and arm gestures are generated (i.e., stage 2) based on the prosodic cues of a new speech-test signal, which will follow the same previously illustrated phases of the training stage.

The main purpose of stage 3 is to setup for a successful long-term human-robot interaction (a future concern for our research), for which the robot should be able to extend

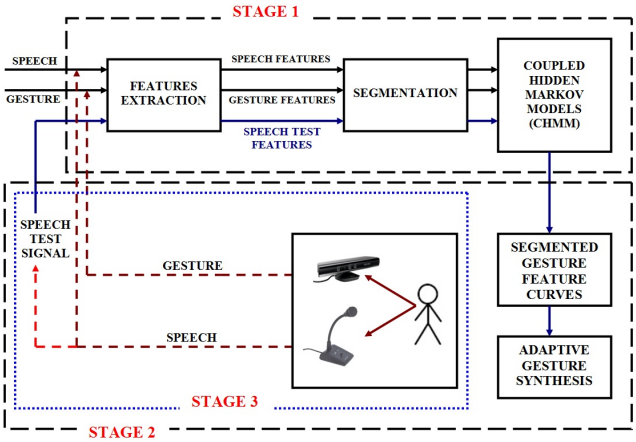

Fig. A1: Overview of the metaphoric gesture generator. More details are available in Aly and Tapus [6]

incrementally the constructed learning database by acquiring more raw speech and gesture data elements from the nearby humans. Therefore, a Kinect sensor should be continuously employed in parallel with the robot in order to precisely calculate the motion curves of articulations, in addition to a microphone to receive the speech signal of a human user. Afterwards, both of the captured prosody and gestures data will undergo the previously explained phases of the training stage 1 so as to increase the robot ability to synthesize more appropriate gestures.

References

[1] A. Aly. *Towards an interactive human-robot relationship: Developing a customized robot's behavior to human's profile*. PhD thesis, ENSTA ParisTech, France, 2014. 12

[2] 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. 6

[3] A. Aly and A. Tapus. 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)*, Bucharest, Romania, 2012. 6

[4] A. Aly and A. Tapus. Prosody-driven robot arm gestures generation in human-robot interaction. In *Proceedings of the 7th ACM/IEEE Human-Robot Interaction Conference (HRI)*, MA, USA, 2012. 6

[5] A. Aly and A. Tapus. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 325–332, Tokyo, Japan, 2013. 2

[6] A. Aly and A. Tapus. Prosody-based adaptive metaphoric head and arm gestures synthesis in human

robot interaction. In *Proceedings of the 16th IEEE International Conference on Advanced Robotics (ICAR)*, pages 1–8, Montevideo, Uruguay, 2013. 4, 6, 7, 12

[7] E. Andre, T. Rist, S.V. 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*, pages 220–255, MA, USA, 2000. MIT Press. 1

[8] J.N. Bailenson and N. Yee. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, 16:814–819, 2005. 3

[9] J.A. Bargh, M. Chen, and L. Burrows. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Personality and Social Psychology*, 71:230–244, 1996. 3

[10] M.R. Barrick and M.K. Mount. The big five personality dimensions and job performance: A meta-analysis. *Personnel Pshychology*, 44:1–26, 1991. 3

[11] G. Beattie and L. Sale. Do metaphoric gestures influence how a message is perceived? The effects of metaphoric gesture-speech matches and mismatches on semantic communication and social judgment. *Semiotica*, 2012(192):77–98, 2012. 10

[12] F.J. Bernieri. Coordinated movement and rapport in teacher-student interactions. *Nonverbal Behavior*, 12: 120–138, 1988. 3

[13] E. Bevacqua, M. Mancini, and C. Pelachaud. Speaking with emotions. In *AISB Convention: Motion, Emotion and Cognition*, Leeds, UK, 2004. University of Leeds. 1

[14] D. Byrne and W. Griffit. Similarity and awareness of similarity of personality characteristic determinants of attraction. *Experimental Research in Personality*, 3: 179–186, 1969. 2

[15] J.N. Cappella and S. Planalp. Talk and silence sequence in informal conversations III: Interspeaker influence. *Human Communication Research*, 7(2):117–132, 1981. 3

[16] J. Cassell and T. Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. In *User Modeling and User-Adapted Interaction*, volume 13, pages 89–132, 2003. 1

[17] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E.F. Churchill, editors, *Embodied Conversational Agents*, pages 29–63. MIT Press, MA, USA, 2000. 1

[18] J. Cassell, H.H. Vilhjálmsson, and T. Bickmore. BEAT: The behavior expression animation toolkit. In *Proceedings of the SIGGRAPH*, pages 477–486, 2001. 4, 5

[19] T.L. Chartrand and J.A. Bargh. The chameleon effect: The perception-behavior link and social interaction. *Personality and Social Psychology*, 76(6):893–910, 1999. 3

[20] 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)*, Barcelona, Spain, 2004. 5

[21] M. Coltheart. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33:497–505, 1981. 4

[22] J.M. Dewaele and A. Furnham. Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49(3):509–544, 1999. 4

[23] N.S. Dicaprio. *Personality theories: A guide to human nature*. Holt, Rinehart and Wilson, NY, USA, 1983. 2

[24] P. Dijkstra and D.P.H. Barelds. Do people know what they want: A similar or complementary partner? *Evolutionary Psychology*, 6(4):595–602, 2008. 2

[25] J. Eriksson, M.J. Matarić, and C. Winstein. Hands-off assistive robotics for post-stroke arm rehabilitation. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 21–24, IL, USA, 2005. 2

[26] H.J. Eysenck. *The structure of human personality*. London, UK, 1953. 2, 10

[27] H.J. Eysenck. Dimensions of personality: 16, 5 or 3? Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12:773–790, 1991. 2, 10

[28] H.J. Eysenck and S.B. Eysenck. *Manual: Eysenck personality inventory*. Educational and Industrial Testing Service, CA, USA, 1968. 10

[29] C. Fellbaum. *WordNet: An electronic lexical database*. MIT Press, MA, USA, 1998. 5

[30] T.W. Finin, A.K. Joshi, and B.L. Webber. Natural language interactions with artificial experts. In *Proceedings of the IEEE,10(2)*, pages 921–938, 1986. 1

[31] K. Forbes-Riley and D.J. Litman. Investigating human tutor responses to student uncertainty for adaptive system development. In *Lecture Notes in Computer Science*, volume 4738, pages 678–689, 2007. 1

[32] K. Forbes-Riley, D.J. Litman, and M. Rotaru. Responding to student uncertainty during computer tutoring: An experimental evaluation. In *Lecture Notes in Computer Science*, volume 5091, pages 60–69, 2008. 1

[33] A. Furnham. *Language and personality*. In Giles, H. and Robinson, W. (Eds.), Handbook of Language and Social Psychology. Winley, 1990. 4

[34] H. Giles and P.F. Powesland. Speech style and social evaluation. In F. Erickson, editor, *Language in Society*, pages 428–433. Cambridge University Press, Cambridge, UK, 1978. 3

[35] A.J. Gill and J. Oberlander. Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 363–368, 2002. 4

[36] L.R. Goldberg. An alternative description of personality: The Big-Five factor structure. *Personality and*

Social Psychology, 59:1216–1229, 1990. 2, 4

[37] L.R. Goldberg. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, 7:7–28, 1999. 2, 4

[38] B.J. Grosz. TEAM: A transportable natural language interface system. In *Proceedings of the Conference on Applied Natural Language Processing*, pages 39–45, CA, USA, 1983. 1

[39] N. Gueguen. *100 petites experiences en psychologie de la seduction*. Dunod, Paris, France, 2007. 3

[40] B.B. Gump and J.A. Kulik. Stress, affiliation, and emotional contagion. *Personality and Social Psychology*, 72: 305–319, 1997. 3

[41] E.T. Hall. *The hidden dimension*. Doubleday, NY, USA, 1966. 3

[42] B. Hartmann, M. Mancini, and C. Pelachaud. Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. In *Proceedings of the Computer Animations*, Geneva, Switzerland, 2002. IEEE Computer Society Press. 1

[43] R. Hassin and Y. Trope. Facing faces: Studies on the cognitive aspects of physiognomy. *Personality and Social Psychology*, 78:837–852, 2000. 4

[44] C.M. Huang and B. Mutlu. Learning-based modeling of multimodal behaviors for humanlike robots. In *Proceedings of the 9th ACM/IEEE Human-Robot Interaction Conference (HRI)*, Germany, 2014. 11

[45] K. Isbister and C. Nass. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *Human-Computer Studies*, 53:251–267, 2000. 2, 4

[46] L.A. J-Campbell, K.A. Gleason, R. Adams, and K.T. Malcolm. Interpersonal conflict, agreeableness, and personality development. *Personality*, 71(6):1059–1085, 2003. 4

[47] C.G. Jung, R.F.C. Hull, and H.G. Baynes. *Psychological types*. Princeton University Press, NJ, USA, 1976. 10

[48] A. Kendon. Gesticulation and speech: Two aspects of the process of utterance. In M.R. Key, editor, *The Relationship of Verbal and Nonverbal Communication*, pages 207–227. Mouton Publishers, The Hague, The Netherlands, 1980. 6

[49] S. Kopp, B. Krenn, S. Marsella, A. Marshall, C. Pelachaud, H. Pirker, K. Thórisson, and H. Vilhjálmsson. Towards a common framework for multimodal generation: The behavior markup language. *Intelligent Virtual Agents*, pages 205–217, 2006. 1

[50] S. Kopp, K. Bergmann, and I. Wachsmuth. Multimodal communication from multimodal thinking - towards an integrated model of speech and gesture production. *Semantic Computing*, 2(1):115–136, 2008. 1

[51] M. Lafrance. Posture mirroring and rapport. In M. Davis, editor, *Interaction Rhythms: Periodicity in Commutative Behavior*, pages 279–298. Human Sciences Press, NY, USA, 1982. 3

[52] B. Lavoie and O. Rambow. A fast and portable realizer for text generation. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP)*, 1997. 5

[53] Q.A. Le and C. Pelachaud. Generating co-speech gestures for the humanoid robot NAO through BML. In *Proceedings of the 9th international conference on Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, pages 228–237, 2012. 1

[54] T.F. Leary. *Interpersonal diagnosis of personality*. Ronald Press, NY, USA, 1957. 2

[55] K.M. Lee, W. Peng, S-A. Jin, and C. Yan. Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human robot interaction. *Communication*, 56:754–772, 2006. 2

[56] K. Leuwerink. A robot with personality: Interacting with a group of humans. In *Proceedings of the 16th Twente Student Conference on IT*, Enschede, The Netherlands, 2012. 3

[57] R.A. Lippa and J.K. Dietz. The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Nonverbal Behavior*, 24:25–43, 2000. 4

[58] L. Lucignano, F. Cutugno, S. Rossi, and A. Finzi. A dialogue system for multimodal human-robot interaction. In *Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI)*, Australia, 2013. 11

[59] P. Luo, V. Ng-Thow-Hing, and M. Neff. An examination of whether people prefer agents whose gestures mimic their own. In *Intelligent Virtual Agents: Lecture Notes in Computer Science*, volume 8108, pages 229–238. 2013. 3

[60] F. Mairesse and M. Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37, 2011. 4, 5

[61] F. Mairesse, M. Walker, M. Mehl, and R. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Artificial Intelligence Research (JAIR)*, pages 457–500, 2007. 4

[62] M. Mancini and C. Pelachaud. Distinctiveness in multimodal behaviors. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 159–166, 2008. 1

[63] R.E. Maurer and J.H. Tindall. Effects of postural congruence on client's perception of counselor empathy. *Counseling Psychology*, 30:158–163, 1983. 3

[64] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, IL, USA, 1992. 1, 10

[65] D. McNeill. *Language and gesture*. Cambridge University Press, Cambridge, UK, 2000. 1, 10

[66] D. McNeill. *Gesture and thought*. University of Chicago Press, IL, USA, 2005. 10

[67] M.R. Mehl, S.D. Gosling, and J.W. Pennebaker. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Personality and Social Psychology*, 90:862–877, 2006. 4

[68] Y. Moon and C. Nass. How real are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, 23:651–674, 1996. 4

[69] L.W. Morris. *Extraversion and introversion: An interactional perspective*. Hemisphere Publishing Corporation, NY, USA, 1979. 2

[70] J.B. Murray. Review of research on the myers-briggs type indicator. *Perceptual and Motor Skills*, 70:1187–1202, 1990. 2

[71] I. Myers-Briggs and P.B. Myers. *Gifts differing: Understanding personality type*. Davies-Black Publishing, CA, USA, 1980. 2

[72] H. Nakajima, S.B.C. Nass, R. Yamada, Y. Morishima, and S. Kawaji. The functionality of human-machine collaboration systems-mind model and social behavior. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, pages 2381–2387, VA, USA, 2003. 3

[73] H. Nakajima, Y. Morishima, R. Yamada, S. Brave, H. Maldonado, C. Nass, and S. Kawaji. Social intelligence in a human-machine collaboration system: Social responses to agents with mind model and personality. *Japanese Society for Artificial Intelligence*, 19(3):184–196, 2004. 3

[74] C. Nass and M.K. Lee. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency attraction. *Experimental Psychology: Applied*, 7:171–181, 2001. 2, 4

[75] V. Ng-Thow-Hing, P. Luo, and S. Okita. Synchronized gesture and speech production for humanoid robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, 2010. 1

[76] R. Niewiadomski, S. Hyniewska, and C. Pelachaud. Modeling emotional expressions as sequences of behaviors. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pages 316–322, Amsterdam, The Netherlands, 2009. 1

[77] E. Park, D. Jin, and A.P. Del-Pobil. The law of attraction in human-robot interaction. *Advanced Robotic Systems*, 9(35), 2012. 2

[78] J.W. Pennebaker and L.A. King. Linguistic styles: Language use as an individual difference. *Personality and Social Psychology*, 77:1296–1312, 1999. 4

[79] J. Pittam. *Voice in social interaction: An interdisciplinary approach*. Sage, CA, USA, 1994. 4

[80] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 257–286, 1989. 12

[81] B. Reeves and C. Nass. The media equation. *University of Chicago Press*, 1996. 2

[82] E. Reiter and R. Dale. *Building natural language generation systems*. Cambridge University Press, Cambridge, UK, 2000. 4

[83] I. Rezek and S.J. Roberts. Estimation of coupled hidden Markov models with application to biosignal interaction modeling. In *Proceedings of the IEEE International Workshop on Neural Networks for Signal Processing (NNSP)*, Sydney, Australia, 2000. 12

[84] I. Rezek, P. Sykacek, and S. Roberts. Coupled hidden Markov models for biosignal interaction modeling. In *Proceedings of the 1st International Conference on Advances in Medical Signal and Information Processing (MEDSIP)*, pages 54–59, UK, 2000. 12

[85] R.E. Riggio and H.S. Friedman. Impression formation: The role of expressive behavior. *Personality and Social Psychology*, 50:421–427, 1986. 4

[86] K.R. Scherer. Language and personality. In K.R. Scherer and H. Giles, editors, *Social Markers in Speech*, pages 147–209. Cambridge University Press, Cambridge, UK, 1979. 4

[87] M. Selfhout, W. Burk, S. Branje, J. Denissen, M.V. Aken, and W. Meeus. Emerging late adolescent friendship networks and Big Five personality traits: A social network approach. *Personality*, 78(2):509–538, 2010. 4

[88] A. Stent, R. Prasad, and M.A. 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)*, pages 79–86, NJ, USA, 2004. 5

[89] H.S. Sullivan. *The interpersonal theory of psychiatry*. Norton, NY, USA, 1953. 2

[90] 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*, CA, USA, 2008. 2

[91] A. Tapus, C. Tapus, and J.M. Matarić. User-robot personality matching and robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics, Special Issue on Multidisciplinary Collaboration for Socially Assistive Robotics*, 1(2):169–183, 2008. 3

[92] R.B. Van-Baaren, R.W. Holland, B. Steenaert, and A. Van-Knippenberg. Mimicry for money: Behavioral consequences of imitation. *Experimental Social Psychology*, 39:393–398, 2003. 3

[93] W.E. Vinacke, K. Shannon, V. Palazzo, and L. Balsavage et al. Similarity and complementarity in intimate couples. *Genetic, Social, and General Psychology Monographs*, 114(1):51–76, 1988. 2

[94] K. Vogel and S. Vogel. L'interlangue et la personnalite de l'apprenant. *Applied Linguistics*, 24(1):48–68, 1986. 4

[95] W. Wahlster and A. Kobsa. User models in dialog

systems. pages 4–34, Germany, 1989. Springer Verlag. 1

[96] J.T. Webb. Interview synchrony: An investigation of two speech rate measures. In A.W. Siegman and B. Pope, editors, *Studies in Dyadic Communication*, pages 115–133. Pergamon Press, NY, USA, 1972. 3

[97] J.L. Williams. Personal space and its relation to extraversion-introversion. *Canadian Journal of Behavioural Science*, 3(2):156–160, 1971. 3

[98] D. Windhouwer. The effects of the task context on the perceived personality of a Nao robot. In *Proceedings of the 16th Twente Student Conference on IT*, Enschede, The Netherlands, 2012. 3

[99] S. Woods, K. Dautenhahn, C. Kaouri, R. Boekhorst, and K.L. Koay. Is this robot like me? Links between human and robot personality traits. In *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Tsukuba, Japan, 2005. 2

[100] S.N. Woods, K. Dautenhahn, C. Kaouri, R.T. 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*, 8(2):281–305, 2007. 2

[101] K. Young. *Source book for social psychology*. A.A. Knopf, NY, USA, 1927. 10

[102] I. Zukerman and D.J. Litman. Natural language processing and user modeling: Synergies and limitations. In *User Modeling and User-Adapted Interaction*, volume 11, pages 129–158, 2001. 1