# The Influence of Voice Pitch on the Evaluation of a Social Robot Receptionist

Andreea Niculescu[1], Betsy van Dijk[2], Anton Nijholt[3]
Human Media Interaction
University of Twente
Enschede, Netherlands
niculescuai@ewi.utwente.nl[1], bvdijk@ewi.utwente.nl[2],
anijholt@ewi.utwente.nl[3]

Swee Lan See
Institute for Infocomm Research
A*STAR
Singapore
slsee@i2r.a-star.edu.sg

*Abstract*—. **In this paper we present an experiment addressing the effect of voice pitch on the evaluation of a social robot receptionist. Twenty eight test participants interacted with two "female" robot characters: one with a high-pitched, exuberant voice, the other with a low-pitched, calm voice. Our results show that the high pitch robot was perceived significantly more attractive in terms of voice, behavior and personality. We also found that the increased level of the robot's attractiveness induced significantly better ratings on the overall enjoyment and overall interaction quality. With our study we would like to stress the importance of the voice, in general (and the voice pitch, in particular) in the social robot design and to encourage further research in this topic within the HCI (in particular HRI – Human Robot Interaction) community.**

*Keywords – social robots, voice pitch, user studies*

## I. INTRODUCTION AND BACKGROUND

The voice is a key element in face-to-face communication not only because it conveys the intended message, but also because it contains highly relevant cues for social interactions. Such cues point to speaker's gender, age, personality, emotional state or place of origin and enable socially intelligent individuals to easily decide who to like, who to trust and who to mate. Therefore, sensitivity to voice and language cues has always played a critical role along evolutionary history in human social groups [1].

The pitch, along with the timbre, volume and speech rate is one of the most important voice characteristics. The pitch refers to how high or low the voice is and is determined by the fundamental frequency. Average pitch values are 120 Hz for male voices and 210 Hz for female voices [2].

Research involving manipulation of vocal characteristics has shown that voice pitch has influence on the perceived attractiveness of a person. The study by Riding et al. [3] found that women rated men with high-pitched voices as to be significantly less attractive than men with medium or low-pitched voices. On the other side, the study by Colin and Missing [4] showed that men found high-pitched female voices more attractive. According to Jones [5] women also show the same preference for high-pitched female voice. Further, voices rated higher in attractiveness are associated with more favourable impressions of overall personality and even with a higher degree of similarity between the test participant and the rated voice [6].

A very recent study [7] associating voice pitch with the retention of content in long term memory found that both high- and low-pitched voices led to better results than medium-pitched voices. This result was independent of whether the speech samples represented natural or manipulated voices.

The pitch is also an indicator for both female and male voices of the person's maturity [8]. A lower pitch, corresponding to a higher maturity was found to be correlated with lower levels of stress and more positive emotions [9]. A higher pitch, on the other hand is associated with more emotional, immature voices indicating greater levels of emotional instability, weakness and psychological tension, in both male and female voices [10]. Impression formation studies investigating the voice alone have found that more mature voices result in impressions of lower warmth and agreeableness [10] but higher dominance and assertiveness [11]-[13].

The pitch form along with volume and the speech rate the four fundamental aspects of voices that indicate personality [22]. Speakers who talks loudly and rapidly in a high pitch using wide voice range variations are more likely to be extrovert while a soft, deep, monotone voice, speaking slowly often belongs to an introvert [1]. Research studies have demonstrated that it is possible to model the personality of synthetic voices using only these four voice aspects [1], [23]; the test subjects [1] not only identified correctly the "personality" behind the synthetic voice but also used this knowledge to guide their feelings and behavior towards the product represented by the voice.

Surprisingly, in the HRI field there is little research done on the psychological effects of the voice, let alone the voice pitch. Walters et al. [14] investigated how people approached a robot when its voice was female, male, highly synthesized (robotic) or when no voice was involved. They found that approached distances were significantly higher when the voice was highly synthesized. Manipulation of affect expression in a robot's voice was investigated in another study [15] in conjunction with task performance. The results showed that voice affect expression can motivate people to perform better on a joint task

when the robot was present in the environment. Apart from these two studies there are no others known to the authors dealing with voice manipulation in HRI.

In our study we investigated the effects of the voice pitch variation on the evaluation of a social robot receptionist. Following the literature findings our hypothesis was that a high-pitched "female" robot would be perceived as generally more attractive, but also more emotional and possibly more extrovert. Additionally, the high pitch robot would achieve higher ratings on the robot's overall personality appeal and degree of perceived similarity with the test participants. Further, we hypothesized that an increased level of attractiveness and personal appeal would lead to a better interaction evaluation in terms of enjoyment, usefulness and overall interaction quality. On the other hand, we would expect the low pitch robot to be perceived as less pleasant but more assertive, stronger, leading to more positive users' feelings and decreasing the interaction stress.

## II. EXPERIMENT DESIGN

For our experiment we used the social robot Olivia. Olivia is the 4th service robot model developed by the A*STAR robotic team from I2R (Singapore). The robot is approximately 152 kg, 1.6 m tall and has 13 degrees of freedom in total: head (3 degrees), body (2 degrees) and hands (2x4 degrees). It is built on a PowerBot base mobile platform and equipped with several hardware/mechanical components, including actuators (servomotor, harmonic gear system, drive unit and harmonic drive servo actuators), a laser (Hokoyu URG-04LX), cameras (Bumblebee®2 and DVN1501 mono camera), microphones and speakers. The robot has several independent software modules for controlling and executing several functions: a motion control (MC), a dialog management system (DMS) and a vision understanding (VU). The MC module employs advance motion control algorithms, such as nonlinear task space control and joint space control to control the robot's movements. The VU module deploys a multi-model fusion maximum likelihood method by integrating four different approaches: stereo-based human detection, HOG-based human detection, color-based tracking, and motion estimation for human detection and tracking. All software modules run two PC boards: one Intel Corei7 (2.8 GHz) and one Atom processor (1.2GHz).

Due to resource limitation the experiment was performed with only one robot. However, to study the effects of different voice pitches two different robot characters are obviously needed. Thus, we symbolically "changed" Olivia's appearance by adding two pigtails and a pink band to her fabric hair (see Fig. 1). We called the new "face" Cynthia. Such simple look modifications, even if minimal are required in order to keep the robot character consistent. Literature studies in the past warned about negative effects of multiple voice personalities associated with the same system (see chapter 5 in [1]). Thus, an experiment design in which we would alternate voice and look a 2x2 matrix should be avoided since it would confuse the participants about the robot's personality inducing unwanted negative effects.



Figure 1. Olivia (left) and Cynthia (right).

To generate the robot's voices the TTS engine from Loquendo was used. The engine allows pitch variations between 0 and 100. A normal inflected voice would have a pitch of 50. Highly animated voices would display pitch values greater than 50 while a pitch value of 0 would produce a flat, monotone voice. For our experiment we chose "Allison", a U.S. English accented voice. We set the pitch value for Cynthia's voice at 20 and for Olivia's voice at 70. All other voice parameters were tuned to the same values: timbre = 50, speech rate = 40, volume = 50. The differences in look and voice were explained as being related to different software modules used for Olivia and Cynthia.

To ensure homogeneity between the trials we substituted the speech recognition module by a human wizard. The prompts were prepared in advance and played back by the wizard during the experiment. The robot was placed in a small observation room designed for usability experiments. Attached to the robot was a screen were information on name, office location, phone number, staff's picture or other building related issues could be displayed. Equipped with headphones and sitting behind a dark screen the wizard team could see and hear the participants interacting with the robot. The experiment was recorded by two cameras placed behind the robot and on the side. The users were informed about the recordings and were asked for permission.

Before entering the room the participants received a short briefing on the experiment proceedings and a scenario describing their tasks: they had to inform the robot about an appointment with a staff member and to ask for direction in the building. Apart from the name of the staff, which was different in the scenarios, the interaction with the robot was designed to be identical in both cases: the robot informed the staff about the visit over the phone and explained the visitor how to get to the staff's office.

The order in which the participants interacted with the robots was randomized to exclude any potential biases that might arise from being exposed to one particular voice pitch before the other.

After each interaction with the robot the participants were given a questionnaire to fill in.

## III. Questionnaire Design

In order to test how the voice pitch affected users' perception of the robot and of the interaction we needed a questionnaire that addressed these two questions.

According to Hassenzahl et al. [16], the user evaluation of a system is influenced by its pragmatic and hedonic quality. Applied to conversational interactions, pragmatic quality would refer to **functional aspects** determining how well a certain communicative goal is achieved, while hedonic quality would relate to **non-functional aspects** indicating how much the user enjoyed the interaction. Looking closer at human robot interaction we can identify the following factors that could shape the human's perceptions about the robot and eventually influence the robot's evaluations in the terms described by Hassenzahl:

- Robot appearance appeal – how the robot presents himself to the outside world, i.e. how he looks, talks, behaves, expresses personality (**hedonic values**).

- Task appeal – how enjoyable (**hedonic value**), effective, efficient and easy it is to interact with the robot (**pragmatic value**).

- Content appeal – how attractive is the content presented (**hedonic value**), how interesting (**hedonic value**) and good (**pragmatic value**) are the answers delivered.

The effect produced by the interplay of functional and non-functional interaction aspects is reflected on the users' feeling during the experiment. Therefore, the questionnaire should also include questions addressing this additional factor.

Since there is no single questionnaire that includes all these factors we designed our own survey using the following questionnaires: AtrakDiff [16] – which evaluates user attitudes towards interactive products, SASSI [17] – which measures the usability of speech dialogue systems and ITU MOS[18] – which determines the voice quality of TTS systems. We also incorporated in our survey items originating from the four Gricean conversational maxims [19] which address basic principles for an effective communication.

The questionnaire was built, similar to AttrakDiff – from which most of the items are taken – using semantic differentials on 7 point bipolar scales.

### A. Robot Appearance Appeal

The appeal of the robot's appearance addresses 4 different aspects comprised in the following subscales: look, voice, speaking style and behavior. Additionally, there is a fifth component – the robot's personality- which includes variables meant to assess different personality dimensions that could not be subsumed in one single subscale.

The robot's look subscale contains 5 items: presentable/unpresentable, professional/unprofessional, pleasant/unpleasant, inviting/rejecting and overall impression (AttrakDiff).

For the robot's voice appeal we used 5 – out of 8 – items from the ITU MOS questionnaire which measures the quality of TTS voices. These are: pronunciation, articulation, speaking rate, pleasantness and overall impression. Two remaining items, listening effort and comprehension problems are referred to on a further subscale addressing the answer quality, while a third one –acceptance- was reformulated into another item referring to the matching degree between the robot's voice and look.

For the third subscale, the speaking style, we used the AttrakDiff word pair stylish/tacky split to two different concepts that are more appropriate for a conversational speech style: polite/impolite and elegant/rough. The subscale also contains a $3^{rd}$ item concerning the overall speaking style impression.

The robot's behavior scale includes mostly word pairs from AttrakDiff questionnaire, such as boring/funny, likeable/disagreeable. The word pair sympathetic/unsympathetic was split in two more precise concepts: friendly/unfriendly and receptive/unreceptive. Additionally, we added another word pair especially relevant for a social robot receptionist: helpful/unhelpful. The scale also contains a statement about the overall robot's behavior impression.

Regarding the robot's personality we included 4 questions inspired from the work of Kahn and De Angeli [20] concerning personality traits dealing with potency, social and emotional competence: extrovert/introvert, rational/emotional, strong/weak and assertive/submissive. The personality 'module' also included an overall statement about overall impression about robot's personality is.

All items related with the robot's appeal subscales hold only hedonic values, meaning that they belong to the non-functional aspect of the interaction.

### B. Task Appeal

The task appeal includes functional (i.e. with pragmatic value) as well as non-functional (i.e. with hedonic value) interaction aspects grouped in two subscales. The non-functional subscale refers to the task enjoyment and subsumes the word pairs: new/common, undemanding/challenging and simple/complicated (AttrakDiff).

The functional aspect refers to interaction features emerging from performing the tasks. The interaction features originated from the SASSI questionnaire and concern: speech accuracy, interaction ease, transparency, speed and flexibility.

### C. Content Appeal

For the non-functional aspects of the content appeal we chose 2 statements concerning the content interest (exciting/lame) and the content presentation (original/conventional). Both word pairs were taken from the AttrakDiff questionnaire.

For the functional aspects we built up a subscale regarding the quality of the content delivered. The subscale includes variable pairs inspired from the Gricean's conversational maxims and concerns the answers' informativeness, trustworthiness, relevance and clarity.

### D. User Feelings

This subscale contains items addressing the user feelings during the interaction with the robot. The following word pairs were included: motivated/unmotivated (AttrakDiff); annoyed

/pleased, tense/calm, confident/insecure, frustrated/satisfied comfortable/awkward, in control/out of control (SASSI).

### E. Overall Judgments and Other Personal Details

The questionnaire also includes global ratings concerning overall enjoyment (**hedonic quality**), overall usefulness (**pragmatic quality**) and overall interaction quality. Additionally, we asked the users about the similarity degree between the robot and themselves and about the robot's capability to socialize. At the questionnaire's beginning the users filled in their details concerning: gender, age, work background, previous experience with social robots and speech recognition applications, expectations regarding the robot and personality type. The personality type was determined via an online Myers Briggs personality test [21]. This particular data was collected with the purpose of determining whether there are interaction effects between users' and robots' (perceived) personality.

## IV. RESULTS AND DISCUSSIONS

A total of 28 people participated in this experiment. One questionnaire could not be considered due to a wizard error: the same voice pitch was played in both trials.

The majority of users (77.80%) were under 30 years while 22.20% were between 31-41 years. 63% were male and 37% were female. Concerning the ethnic groups the majority (81.48%) were Chinese, 7.40% Malay, 3.70% Indian and 7.40% European/North American.

44% of the participants had no knowledge of social robots, 48.10% had seen and/or read about them. A small percentage (7.40%) had interacted with a social robot while an even smaller part (3.40 %) was involved in robotic design.

66.70% of the participants had no experience with speech recognition systems. 33.30% indicated having some occasional experience with speech recognition built-in applications for mobile phones, PC software, social robots or web interfaces. The use was rather seldom; only one person reported a frequent use. However, half of the people who used speech recognition devices did it for real purposes, such as airport enquiries, database searches, ordering a pizza or interacting with speech handicapped people. The other half did it for testing or hobby purposes. Four people reported having had good experience with such devices, one was neutral and 3 reported negative experiences.

46.10% of the participants had high expectations, 19.20% had standard or medium expectations while 34.55% indicated to have low expectations.

The test users expected the robot to be able to recognize speech and gestures (21.72%), and to respond well (26.05%) and fast (4.34%) to standard questions. The robot was expected to speak clearly and to express a proper body language (4.34%). While some participants expected the robot to be somewhat restricted to basic tasks and having limited capabilities (17.37%), others expected the robot to act as in real life (4.34%) having a friendly disposition (4.34%), a nice voice tone (4.34%) and even to be able to understand the feelings of the users (4.34%).

Concerning the users' personality profiles 40% of the participants were extrovert while 55.60% were introvert.

### A. Scales Reliability

Firstly we performed a reliability analysis on our scales to verify the internal consistency of the items. All our scales show relatively high Cronbach' $\alpha$ values: between $\alpha = 0.731$ and $\alpha = 0.903$.

Table I presents the Cronbach's $\alpha$ values for the robot's look ($\alpha = 0.856$), robot's voice ($\alpha = 0.801$), robot's speaking style ($\alpha = 0.782$) and robot's behavior appeal ($\alpha = 0.903$).

Only one variable (professional look) was removed from the robot's look subscale, which shows a low correlation ($\alpha = 0.214$) with the other subscale items.

The subscales addressing the task and content appeal, as well as the user's feelings also show high internal consistency: $\alpha = 0.731$ for interaction features, $\alpha = 0.795$ for content quality, $\alpha = 0.856$ for task appeal and $\alpha = 0.846$ for user feelings (see Table II).

TABLE I. ROBOT'S APPEAL ON LOOK, VOICE, SPEAKING STYLE AND BEHAVIOR SUBSCALES

| Look | Items | Item-Total Correlation r | Speaking Style | Items | Item-Total Correlation r |
|---|---|---|---|---|---|
| Items: 4 $\alpha = 0.856$ | 1.Presentable | 0.717 | Items: 3 $\alpha = 0.782$ | 1. Elegance | 0.680 |
| | 2. Pleasant | 0.741 | | 2. Politeness | 0.611 |
| | 3. Inviting | 0.557 | | 3.Overall impression | 0.597 |
| | 4. Overall impression | 0.797 | | | |
| *Removed* | *Professional* | *0.214* | | | |
| **Voice** | **Items** | **Item-Total Correlation r** | **Behavior** | **Items** | **Item-Total Correlation r** |
| Items: 6 $\alpha = 0.801$ | 1. Articulation | 0.544 | Items: 6 $\alpha = 0.903$ | 1. Boring | 0.701 |
| | 2. Pronunciation | 0.568 | | 2. Friendly | 0.782 |
| | 3. Match look | 0.488 | | 3. Likeable | 0.843 |
| | 4. Speaking rate | 0.728 | | 4. Receptive | 0.668 |
| | 5. Pleasantness | 0.407 | | 5. Helpful | 0.681 |
| | 6. Overall impression | 0.642 | | 6. Overall impression | 0.784 |

TABLE II. CONTENT QUALITY, INTERACTION FEATURES, TASKS AND USER FEELINGS SUBSCALES

| Content Quality | Items | Item-Total Correlation r | Tasks | Items | Item-Total Correlation r |
|---|---|---|---|---|---|
| Items: 4 $\alpha = 0.795$ | 1. Thrustworthy | 0.471 | Items: 3 $\alpha = 0.856$ | 1.Common | 0.641 |
| | 2. Informative | 0.715 | | 2. Undemanding | 0.888 |
| | 3. Relevance | 0.617 | | 3. Simple | 0.889 |
| | 4. Clarity | 0.635 | | | |
| **Interaction Features** | **Items** | **Item-Total Correlation r** | **User Feelings** | **Items** | **Item-Total Correlation r** |
| Items: 5 $\alpha = 0.731$ | 1. Speech | 0.487 | Items: 7 $\alpha = 0.846$ | 1. Motivated | 0.575 |
| | 2. Transparency | 0.464 | | 2. Pleased | 0.687 |
| | 3. Ease | 0.430 | | 3. Calm | 0.606 |
| | 4. Fastness | 0.477 | | 4. Confident | 0.645 |
| | 5. Flexibility | 0.595 | | 5. Comfortable | 0.662 |
| | | | | 6. Satisfied | 0.632 |
| | | | | 7. In control | 0.451 |

## B. Evaluation Results

To detect differences in ratings between the two robots' characters we performed a Wilcoxon signed rank test. The results show Olivia was perceived as having significantly more appeal in terms of voice, ($p = 0.038$). The test participants also found Olivia to have a more appealing behavior ($p < 0.000$) and a more pleasant overall personality, ($p = 0.029$), thus to be more agreeable than Cynthia. As expected Olivia also appeared to be a more extrovert robot ($p = 0.031$) and users found her more emotional ($p = 0.001$) and much more like themselves ($p = 0.009$). On the other hand Cynthia was perceived as being stronger ($p = 0.046$), more rational ($p = 0.001$) but, contrary to our prediction not more assertive. Also, Olivia was the one inducing more positive feelings in users ($p = 0.027$), and not Cynthia as predicted.

The interaction features subscale was better rated in the "Olivia" condition ($p = 0.044$). The content presented by the high pitch robot – even thought identical (apart from the staff's name) with the one presented by its low pitch counterpart – was perceived as more exciting ($p = 0.002$). The interaction with Olivia was considered much more enjoyable ($p = 0.007$, **hedonic quality**), but not significantly more useful (**pragmatic quality**). On the other side, the overall interaction quality with Olivia was higher ranked ($p = 0.001$) as we expected.

No significant differences could be found for the content presentation or for the subscales concerning speaking style, tasks appeal, answer quality or robot's look – the last one most probably due to the minimal difference between Olivia and Cynthia's look.

Next, we checked the effects of gender, age, expectations, personality, work background, previous experience on the evaluation using a Mann-Whitney test.

Male participants rated the overall personality appeal ($p = 0.013$) of the high pitch robot much better than the female participants. They also indicated significantly higher values for positive feelings towards the robot ($p = 0.029$) and overall enjoyment ($p = 0.013$). They also found the interaction more transparent ($p = 0.036$). On the other hand, female participants assessed Olivia's personality as being more extrovert than male participants ($p = 0.006$) did. Male participants also found Cynthia's look significantly more professional ($p < 0.000$), her voice matched the face better ($p = 0.039$) and her speech was faster ($p = 0.025$). The content presented by Cynthia appeared to them significantly less lame ($p = 0.008$) and her overall behavior more appealing ($p = 0.045$). The men's higher ratings for both robots might be explained by a general stereotype tendency which predicts men's preference for female robots, also found by other studies [24]. The gender difference in rating the overall impression of the robot's personality and behavior would also support the findings reported in [25] showing that men compared with women would more readily treat a robot as social entity.

There were no effects found of age, work or knowledge background in the robotic field. However, some interesting effects were found for those who had no experience with speech recognition devices: both robots were perceived as being significantly more sociable ($p = 0.012$ for Olivia and $p = 0.002$ for Cynthia) by those who had no experience with speech recognition devices. This finding suggests that more experienced users tend to be less open to perceive the robot as a social presence.

Users with higher expectations towards the robot found Olivia to be more flexible ($p = 0.036$). On the other hand users with low expectations found Cynthia's speaking style ($p = 0.023$) and answer quality better ($p = 0.036$). This outcome might be explained by the fact that the majority (71%) of the users with high expectations kept their question strictly limited on the scenario settings while almost half of the users with low expectations (48%) asked some additional questions. Most "unexpected" questions caused some delay – since not all could be prepared in advance by the wizard team - and might have caused the impression of an inflexible answer behavior. On the other hand, the answers improved the information richness and therefore, the answer quality could have been perceived as being better. As for the speaking style we have no explanation, except maybe for the fact that the test participants who interacted longer with the robot could have been more aware of its polite and elegant style of speaking.

Finally, introvert users found the interaction with Cynthia much easier ($p = 0.012$) and her behavior significantly more receptive ($p < 0.000$) than extrovert users. Since Cynthia was perceived as being a more introvert robot, our result is conform to the findings of Revees and Nass [26] and Nass and Brave [1] concerning the attraction effects between human and computers and/or voice with similar personalities.

## V. CONCLUSION

The main focus of this study was to analyze the relationships between the voice pitch and the way people perceive and rate a social robot. Our results show that users seem to prefer high-pitched robot voices since the manipulation had a high impact on robot's attractiveness which further influenced the evaluation results of the entire interaction. Olivia, the higher pitched robot was better rated in terms of overall appearance, behavior and personality; users also felt more positive feelings towards Olivia and rated the interaction concerning both hedonic and pragmatic quality better, as compared to Cynthia.

Therefore, choosing the right voice pitch should be a priority in social robot design. With our study we would like to stress the enormous importance of the voice in human robot interaction and to encourage further research on this topic.

REFERENCES

[1] C. Nass C. and B. Reeves. *Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship*, Cambridge, MIT Press, 2005.

[2] H. Traunmüller and A. Eriksson. (1994) *The Frequency Range of the Voice Fundamental in the Speech of Male and Female Adults*, Manuscript, Dep. Of Linguistics, University of Stockholm. [Online]. Available: http://www.ling.su.se/staff/hartmut/aktupub.htm

[3] D. Riding, D. Lonsdale, and B. Brown. "The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women," *Journal of Nonverbal Behavior*, 30, 2006, pp. 55—61

[4] S. A. Collins and C. Missing. "Vocal and visual attractiveness are related in women,"*Animal Behavior*, 65, 2003, pp. 997-1004

[5] B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovic. "Integrating cues of social interest and voice pitch in men's preferences for women's voices," *Biology Letters*, 4, 2008, pp. 192—194

[6] M. Zuckerman and K. Miyake. "The attractive voice: What makes it so?," *Journal of Nonverbal Behavior*, 17(2), 1993, pp.119—135

[7] H. Helfrich, and P. Weidenbecher, „Impact of voice pitch on text memory," *Swiss Journal of Psychology*, Vol 70(2), 2011, pp. 85—93.

[8] J. M. Montepare and L. Zebrowitz-McArthur. "Perceptions of adults with childlike voices in two cultures," *Journal* of *Experimental Social Psychology*, 23, 1987, pp. 331—349

[9] G. Fairbanks, "Recent experimental investigations of vocal pitch in speech," in G. Fairbanks, *Experimental Phonetics: Selected Articles* Urbana and London: University of Illinois Press, 1966, pp. 161--170.

[10] K.R Scherer, "Personality markers in speech," in K. R. Scherer and H. Giles, (Eds.), *Social Markers in Speech* New York: Cambridge University Press, 1979, pp. 147--209

[11] M. Zuckerman, K. Miyake, and C. S Elkin. "Effects of attractiveness and maturity of face and voice on interpersonal impressions," *Journal of Research in Personality,* 29, 1995, pp. 253—272

[12] D. R. Feinberg, B. C. Jones, A. C. Little, D. M. Burt, and D. I. Perret. "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices," *Animal Behaviour*, 69, 2005, pp. 561--68

[13] D. A. Puts, S. J. C. Gaulin, and K Verdolini, "Dominance and the evolution of sexual dimorphism in human voice pitch," *Evolution and Human Behavior*, 27(4), 2006, pp. 283—296

[14] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. te Boekhorst, "Human approach distances to a mechanical-looking robot with different robot voice styles," in *17th IEEE* International Symposium on Robot and Human Interactive Communication (*RO-MAN*), München, Germany, August 2008, pp.707--712

[15] M. Scheutz and P. Schermerhorn. "Affective goal and task selection for social robots," in Jordi Vallverd and David Casacuberta, (Eds.) *The Handbook of Research on Synthetic Emotions and Sociable Robotics.* IGI Global, 2009

[16] M. Hassenzahl, M. Burmester, and F. Koller, "AttrakDiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in: J. Ziegler and G. Szwillus (Hrsg.), *Mensch and Computer*, Interaktion in Bewegung, Stuttgart, Leipzig: B.G. Teubner, 2003, pp.187—196

[17] K. S. Hone and R. Graham, "Towards a tool for the subjective assessment of speech system interfaces (SASSI)," *Natural Language Engineering*, 6 (3-4), 2000, pp. 287-303

[18] ITU-T Recommendation, "Telephone transmission quality subjective opinion tests. A method for subjective performance assessment of the quality of speech voice output devices," 1994, pp. 85

[19] P. Grice, "Logic and conversation," in P. Cole and J. Morgan, (Eds.) *Syntax and Semantics,* vol 3, New York: Academic Press, 1972

[20] Khan, R. and A. De Angeli, "The Attractiveness Stereotype in the Evaluation of Embodied Conversational Agents", in *INTERACT* (1), 2009, pp. 85-97

[21] Myers Briggs *Online Personality Test.* Available: http://www.humanmetrics.com/cgi-win/jtypes2.asp

[22] W. L. A. Apple, Streeter, and R. M. Krauss, "Effects of pitch and speech rate on personal attributions," *Journal of Personality and Social Psychology*, 37, 1979, pp. 715-727

[23] J. Trouvain, S. Schmidt, M. Schroeder, M. Schmitz, and W.J. Barry", Modelling personality features by changing prosody in synthetic speech," in *Speech Prosody*, 2006

[24] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive robotics: The influence of robot gender on human behavior," in *IROS,* 2009*,* pp. 2563-2568

[25] P. Schermerhorn, M. Scheutz, and C. R. Crowell, "Robot social presence and gender: Do females view robots differently than males?" in *3rd ACM/IEEE International Conference on Human Robot Interaction,* Amsterdam, The Netherlands: ACM, 2008, pp. 263–270

[26] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press/CSLI, New York.1997