Social evaluation of artificial agents by language varieties

Brigitte Krenn, Stephanie Schreitter, Friedrich Neubarth, and Gregor Sieber

Austrian Research Institute for Artificial Intelligence, 1010 Vienna, Austria,
firstname.lastname@ofai.at

Abstract. In Sociolinguistics, language attitude studies based on natural voices have provided evidence that human listeners socially assess and evaluate their communication partners according to the language variety they use. Similarly, research on intelligent agents has demonstrated that the degree an artificial entity resembles a human correlates with the likelihood that the entity will evoke social and psychological processes in humans. Taking the two findings together, we hypothesize that synthetically generated language varieties have social effects similar to those reported from language attitude studies on natural speech. We present results from a language-attitude study based on three synthetic varieties of Austrian German. Our results on synthetic speech are in accordance with previous findings from natural speech. In addition, we show that language variety together with voice quality of the synthesized speech bring about attributions of different social aspects and stereotypes and influence the attitudes of the listeners toward the artificial speakers.

Keywords: language-attitude study, synthetic voices, virtual character design, social evaluation

1 Introduction

In the present paper, we explore the question in how far the language variety used by an artificial agent influences how the human communication partner socially perceives and evaluates the agent. Based on results from language attitude studies on natural voices, we predict that synthetic voices representing a standard language variety are rated differently than colloquial or dialectal synthetic voices. Additionally, we investigate effects of synthesized standard and non-standard language varieties with respect to those aspects of a character relevant for the social interpretation and evaluation of the character. This may influence how the human assesses the intelligence, naturalness, politeness, etc. of the character, and related assumptions such as education, profession, social background of the character, its habits and preferences. A better understanding of such effects is crucial for the design of artificial human-like agents, where the agent's voice, its appearance and behaviour, as well as the application context must match.

In order to gain deeper insights into aspects of social interpretation that are triggered by language variety, we conducted a language attitude study on

three Austrian language varieties, a standard Austrian German male voice, a colloquial Viennese female voice, and a dialectal Viennese male voice. To the best of our knowledge, this kind of language attitude studies on synthesized language varieties are novel. The synthetic voices are implemented with the open domain unit selection speech synthesis engine Multisyn of Festival.[1] For further details on speaker selection and the design of the voices see [22].

To set a context for the language varieties, we have built the synthetic voices into an existing cultural heritage application, where an invisible tourist guide (represented by voice only) accompanies a human visitor within an interactive 3D model of the Baroque State Hall of the Austrian National Library which is one of the world's finest historic libraries. Visitors are able to seek their own interactive ways through the State Hall and ask the guide for information or they may follow the guide on selected virtual tours. For further details on the application see [11, 27]. For the language-attitude study, we have replaced the existing voice within the application with the three Austrian varieties. A part of the guided tour covering the statues in the State Hall was used as material for assessing human evaluations of the synthesized language varieties. We chose this application because it provides a context for the appearance of a disembodied/invisible agent and thus allows concentration on characteristics conveyed by language variety, ensuring that social interpretation of the tour guides solely arises from language variety and voice quality.

In the following sections, we first introduce related work (Sec. 2). We then present the design of the experiment including hypotheses, methods employed, and characteristics of the group of participants (Sec. 3). This is followed by an analysis and discussion of the data (Sec. 4). A summary and outlook is presented in Sec. 5.

2 Related Work

Since the 1980s, research has been carried out showing that the degree an artificial entity resembles a human correlates with the likelihood that the entity will evoke social and psychological processes in humans, e.g. [32, 31]. In psychology, attribution theory focuses on interpretations and ascription of causality to events by individuals, see [7, 10, 30]. Dubinsky [5] summarizes the key assumptions of these slightly different approaches as follows: 1) people try to determine causes of their own behaviour and the behaviour of others; 2) people assign causal explanation for behaviour in a systematic manner; and 3) attributions people make have consequences for future behaviour.

The social categories humans belong to are often activated automatically. Rakic et al. [24] investigated social categorization by using auditory stimuli such as accents and visual stimuli such as looks either separately or in combination to indicate ethnicity. The results showed a similar degree of ethnic categorization by accents and looks, although there was a clear predominance of accents as

[1] See Black & Clark, The Festival Speech Synthesis System, [3], `http://www.cstr.ed.ac.uk/projects/festival/`.

meaningful cues for categorization when the two ethnic cues of looks and accents were combined by creating cross categories. Humans also apply social rules in human-computer interaction normally reserved for interactions with other humans, e.g. [19, 18]. The use of language and producing human-sounding voice are aspects where computers appear specifically human [14]. Social responses towards computational artefacts may be intentionally designed by their creators, but they often affect users in ways that were not foreseen by their developers, e.g. stereotypical reactions towards male and female artificial agents [20].

Therefore, a better understanding of gender, language variety, social and ethnicity effects on users are of crucial importance in order to develop personalized companions accompanying and supporting users. In several experiments on gender-specific effects of language-based systems gender-specific embodiment as well as the voice of an agent have strong impact on the human perception of and preferences for the agent, for instance by Nass and Brave [17]. The results of their experiments support findings in the field of gender linguistics, including that social identification and proximity to communication partners of the same sex is higher than to ones of opposite sex, and that male agents tend to be rated as more competent by both men and women. Crowell et al. [4] conducted an experiment comparing sex-related differences in reactions towards gendered synthetic voices that are either physically embodied within a robot or disembodied. Both men and women found the disembodied female voice and the male embodied voice to be more reliable. Concerning ethnic identity, Cassell and co-workers have studied how children evaluate effects of verbal and non-verbal behaviour of their virtual peers [9, 2]. In the context of human-robot interaction, there is evidence that women tend to rate both male and female synthetic voices more positively than men do [21, 26]. The same effect is found when subjects evaluate natural human voices [29, 28].

In general, speakers of a standard language variety or other varieties attributed with prestige get higher evaluations for competence-related scores, i.e. intelligence, education etc. than non-standard speakers, see [29, 15, 16, 13, 25], and unlike other Austrian dialects, the Viennese dialect is perceived as characteristic for a lower social class [16, 22]. Upper, middle and lower class respondents in Vienna do not attach prestige to dialect usage, cf. [15]. In Moosmüller's study, speakers of Viennese dialect were rated as not very intelligent, tolerant, kind-hearted, friendly, likable or honest. Findings on dialect usage in Linz (Upper Austria), on the contrary, have shown higher social acceptance and appreciation of the local dialect [29]. Soukup states that, other than in Vienna, in Linz and surroundings dialect is spoken in official and public situations much more regularly.

3 Hypotheses, Methods and Participants

3.1 Hypotheses

Attitudes towards natural language variants have been widely examined. But what about synthetic voices? In the present study, quantitative and qualitative

methods are applied: A semantic differential is employed to investigate possible differences between the three synthetic language varieties. It is complemented with open questions to uncover further assumptions of the human listeners about the presumed characters behind the voices, generating hypotheses about living situation of the artificial agent, attribution of social class, age, etc. The quantitative part of the study allows the data on human evaluation of synthesized language varieties to be compared with existing results from language attitude studies based on natural voices. The qualitative part reveals additional insights related to concrete attributions based on language variety and voice quality.

In the following, test hypotheses are presented for evaluating the data collected by means of the semantic differential. The developed hypotheses based on the open questions will be presented in Section 4.2. The main hypothesis of the presented study is that synthetic variants of Austrian German have similar social effects on people living in Austria with German as their mother tongue as natural variants of Austrian German have. Referring to previous work on social evaluation of Austrian language varieties based on natural speech and on the evaluation of synthetic speech in the context of artificial agents as introduced in Section 2, we formulate the following hypotheses to be tested:
H1: The synthetic standard Austrian variant is evaluated as more intelligent, competent, educated and refined than the synthetic dialectal variant.
H2: The synthetic dialectal variant is evaluated as more open-minded, relaxed, natural and with a higher sense of humour than the synthetic standard Austrian variant. See [29, 15, 16, 13, 25] for evidence of H1 and H2 regarding natural language varieties.
H3: Male and female subjects differ in their evaluation of the synthetic language varieties.
H3a: Female subjects consistently rate the different synthetic language variants higher than male subjects do. See [21, 26] for evidence of H3 and H3a from human-robot-interaction and [28, 29] from natural language varieties.

3.2 Methods

Up to date, the most commonly applied method for speaker evaluation is a variant of the matched guise technique, originally developed by Lambert et al. [13, 12]. In the original version, one speaker recites the same text in different language varieties which are then rated by listeners. In the adapted version of the matched guise technique, different speakers are evaluated for different language varieties, e.g. [28, 6]. This adapted version is also used in the present study.

First, the text containing information about the statues in the State Hall of the Austrian National Library was read to the participants by the facilitator. This was followed by the presentation of three videos, each of which showing the same guided tour with the text previously read by the facilitator being synthesised – each video featuring a different synthetic language variety (Austrian Standard German, colloquial and dialectal Viennese). After the subjects had watched the three videos in a row, they had to fill in a questionnaire rating the three invisible tourist guides on a 5-point bipolar semantic differential covering

19 adjective pairs such as 'likable' - 'unlikable', 'educated' - 'uneducated' , 'trustworthy' - 'untrustworthy', etc. See the x-axis of Figure 1 for the list of adjectives (positive poles) employed. The adjective pairs and rating dimensions reflect past research on language attitudes in various contexts, cf. [12, 33, 28], thus allowing the novel results gained for synthetic speech in the present study to be compared with existing results from natural speech. To counteract habituation, the order of the presentation of the adjective pairs (positive, i.e. socially more desirable, and negative, i.e. socially less desirable, poles) in the semantic differential was varied. See Table 1 as an example of adjective pairs in the semantic differential presented in the questionnaire.

sympathisch (likable)						unsympathisch (dislikable)
gebildet (educated)						ungebildet (uneducated)
nicht vertrauenswürdig (not trustworthy)						vertrauenswürdig (trustworthy)

Table 1. Sample adjective pairs from the semantic differential.

Additionally, the subjects responded to a set of open questions including questions regarding their assumptions concerning the characters behind the voices, their assessment of the individual language varieties in the cultural-historic context of the application, as well as their general assessment of Viennese language and people.

3.3 Participants

The study was conducted during two interdisciplinary lectures at the University of Vienna visited by students of Medicine, Cognitive Science, Journalism and German Philology, a lecture at the Technical University Vienna and two lectures at Technikum Wien which is a Technical University of Applied Sciences. In the following, we briefly discuss characteristics of the group of participants.

Sample: 91 Austrian and German students participated in the study, 54 of which were female, 37 male. The subjects were between 18 and 26 years old, except for two who were 31 and 46.

Mother tongue and language use: As the participants of our study are students in Vienna, the vast majority (75 of a total of 91) of which also resides in the city. 34 of the participants spent their youth in Vienna, 20 in Lower Austria, and 11 in Germany. The rest comes from other provinces of Austria, one participant originates from South Tyrol. All participants have German mother tongue, 28 of them state that they use dialect or colloquial variety, another 18 only use standard language, whereas 43 state that they use both dialect/colloquial and standard language. Approximately half of those 18 who state that they

use standard language only either come from Germany (6) or have at least one non-German-speaking parent (4).

4 Data Analysis and Discussion of Results

4.1 Comparing the Language Varieties

In order to test for differences between the language varieties represented by TG1, TG2 and TG3, pairwise Wilcoxon tests have been conducted. Thus, we account for the nonparametric nature of the data and their ordinal scale (semantic differential with 5-point Likert scale ratings). The Bonferroni correction was applied to counteract the problem of multiple comparisons. Bonferroni is only one, simple means to correct for multiple comparisons which in general is a broadly and controversially discussed topic. See for instance [1] for a start. As the Bonferroni correction is a rather conservative means to account for type 1 error, we also applied Holm's sequential Bonferroni which is less conservative than original Bonferroni [8]. Overall, there was only one p-value additionally becoming significant when sequential Bonferroni was applied, namely when comparing TG1 and TG3 on the dimension *not arrogant*. For all other comparisons the significances remain the same as with the original Bonferroni correction. Thus we base our interpretation on the results gained from applying the Bonferroni correction. In Table 2, the results are listed for the pairwise comparisons of the language variants along the adjective dimensions: ns stands for not significant and s for significant; p-values below $\alpha = 0.05$ are indicated with * and p-values below $\alpha = 0.01$ are indicated with **. Results per comparison are presented in the following order: significance level according to Wilcoxon test (α); significance level according to Bonferroni correction (α_{Bonf}); p-values resulting from the Wilcoxon tests.

Results: For all three comparisons TG1 versus TG2, TG1 versus TG3, TG2 versus TG3, applying Bonferroni correction, no difference in human evaluation of the language varieties are found in the dimensions *likability friendliness* and *arrogance*. Whereas all three varieties mutually differ with respect to the dimensions *educated, sense of humour, serious*. With TG1, the standard speaker, being perceived as more educated and serious than TG2, the speaker of colloquial Viennese, and both TG1 and TG2 being perceived as more educated and serious than TG3, the speaker of Viennese dialect. Regarding *sense of humour*, the order is reversed, with TG3 being perceived as having the highest sense of humour and TG1 the least.

Overall, TG1 and TG2 are evaluated differently in 7 out of 19 dimensions, TG2 and TG3 in 12 out of 19, and TG1 and TG3 in 16 out of 19. In other words there are many significant differences in the evaluation of the Austrian standard (TG1) and the Viennese dialect (TG3), whereas TG1 and TG2 are more closely related as of how the variants are perceived by the human listeners. In particular: TG1, TG2, TG3 decrease in educatedness and seriousness (i.e., TG1 is perceived as being significantly more educated and serious than TG2 and

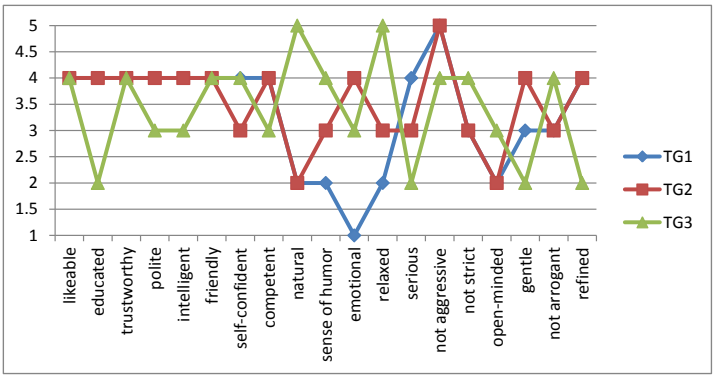

Fig. 1. Line diagram of median scores for each speaker. 5 indicates 'very likable', 'educated', 'trustworthy' etc.; 1 indicates 'unlikable', 'uneducated', 'not trustworthy' etc. TG1 (male voice, Austrian German), TG2 (female voice, colloquial Viennese), TG3 (male voice, Viennese dialect)

Results Wilcoxon Test – Significance Levels									
	TG1 vs TG2			TG1 vs TG3			TG2 vs TG3		
	α	α_{Bonf}	p	α	α_{Bonf}	p	α	α_{Bonf}	p
likable	s*	ns	0.033	ns	ns	0.280	ns	ns	0.159
educated	s**	s**	0.000	s**	s**	0.000	s**	s**	0.000
trustworthy	s**	s*	0.001	s**	s**	0.000	ns	ns	0.064
polite	s*	ns	0.032	s**	s**	0.000	s**	s**	0.000
intelligent	s*	ns	0.038	s**	s**	0.000	s**	s*	0.000
friendly	ns	ns	0.711	ns	ns	0.159	ns	ns	0.128
self confident	s**	ns	0.007	s**	s**	0.000	s**	s**	0.000
competent	s**	s**	0.000	s**	s**	0.000	s*	ns	0.025
natural	s*	ns	0.033	s**	s**	0.000	s**	s**	0.000
sense of humour	s**	s**	0.000	s**	s**	0.000	s**	s**	0.000
emotional	s**	s**	0.000	s**	s**	0.000	ns	ns	0.992
relaxed	ns	ns	0.150	s**	s**	0.000	s**	s**	0.000
serious	s**	s**	0.000	s**	s**	0.000	s**	s**	0.000
not aggressive	ns	ns	0.496	s**	s**	0.000	s**	s**	0.000
not strict	ns	ns	0.126	s**	s**	0.000	s*	ns	0.047
open minded	ns	ns	0.470	s**	s*	0.001	s**	s**	0.000
gentle	ns	ns	0.072	s**	s**	0.000	s**	s**	0.000
not arrogant	ns	ns	0.983	s*	ns	0.016	s*	ns	0.038
refined	s*	ns	0.036	s**	s**	0.000	s**	s**	0.000

Table 2. Pairwise comparison of language variants: levels of significance according to Wilcoxon (α) and with Bonferroni correction (α_{Bonf}), p-values; ns not significant, s significant, * $\alpha = 0.05$, ** $\alpha = 0.01$.

TG2 as being significantly more educated than TG3). TG1 is more trustworthy and competent than TG2 and TG3. TG1 and TG2 are perceived as more polite and intelligent than TG3. TG3 is more self-confident, natural, relaxed, open minded but also more aggressive, and less gentle and refined than TG1 and TG2. There is a decrease of perceived sense of humour from TG3 to TG2 to TG1, with TG1 ranking lowest in sense of humour. TG3 and TG2 are perceived as more emotional than TG1. TG3 is perceived as less strict than TG1. The line diagram in Figure 1 illustrates the evaluation patterns related to the three language variants.

Summing up, the results from comparing TG1 and TG3 support our hypotheses H1 and H2, and are comparable to results based on natural speech where the standard language variant is perceived as more educated, serious, intelligent, trustworthy, competent and polite than the dialectal variant, whereas the dialectal variant is perceived of having a bigger sense of humour, being more relaxed and emotional but also being more aggressive.

Mann-Whithey-U tests were applied to test for H3, cross-classifying sex as independent variable with TG1, TG2 and TG3. Without correcting for repeated comparisons, differences in the evaluations by men and women are found for polite (TG1), gentle (TG1, TG2), natural, non strict, open minded, intelligent (TG3), with females rating more positively than males, except for intelligence of TG3 (females rate TG3 less intelligent than males do). Applying sequential Bonferroni as well as Bonferroni, only one difference remains significant, namely TG3 on the dimension *strict - non strict* – females found TG3, the Viennese dialectal variant, less strict than males did. With the rejection of H3, also H3a is rejected, providing evidence that sex difference may be negligible with respect to our data.

4.2 Social Interpretation and Evaluation

Further, the participants responded to a set of open questions, assessing each artificial speaker's typicalities, and providing general comments on standard and dialectal language use. With the open questions we aimed at exploring the listeners' believes about the speakers, in particular: a) whether there are specific tendencies of belief; b) if and how these assumptions differ between the voices and related language varieties; c) how the respective varieties fit the application scenario; d) how Viennese language varieties and Viennese people are evaluated by the subjects.

Social interpretation: The participants were asked how they imagine the respective character behind the voices of TG1, TG2 and TG3, respectively. *Eine Person, die wie der erste | zweite | dritte Tourguide spricht, stelle ich mir folgendermassen vor...* (A person who speaks like the 1st | 2nd | 3rd tour guide, I imagine to be ...)

Assumptions about where the person behind the voice lives and what profession she or he might have were most prominent for all three characters. Answers to the open question were given by almost all subjects (87 for TG1 and 89 for

TG2 and TG3, respectively, out of 91 subjects total). The prevalent characteristics attributed to the different voices are:

TG1 *lives in the city* (26 mentions, 9 of which say Vienna), is a *professional speaker* (22 mentions) and an *academic* (19 mentions).
TG2 is an *elderly person* (62 mentions), *retired* (33 mentions) and *lives in the city* (16 mentions, 9 of which say Vienna).
TG3 *lives in the city* (29 mentions, all say Vienna), *lives at the country* (26 mentions), *likes to go to the pub* (20 mentions), *is a peasant* (10 mentions).

These results show, on the one hand, a perceived connection between synthesised Viennese language variety and place of residence, with increasing percentage of mentions of Vienna, the more dialectal the voice appears. On the other hand, they show a perceived connection between dialect and rural origin which has been attested also in previous research, e.g. [16]. As regards other factors, 22 participants expressed the opinion that TG1 works for broadcast media. In [29], the standard speaker was also believed to work in public media. Additionally, 19 participants believe TG1 is an academic. A majority of subjects agree that TG2 is an elderly (62) or retired (33) person. For TG3, 20 participants speculate that he likes to go to the pub. Thus we see a clear distinction between the character-specific interpretations that are triggered by the voices. While for TG1 the characteristics of being an urbanite, a professional speaker and academic are most prominent, it is age for TG2, and regional attribution (Vienna) as well as a marked preference for going to the pub for TG3. In other words, not only language variety but also other vocal characteristics such as age are relevant for social interpretation and attribution.

Local variety and cultural context: Referring to the question which of the tour guides is the preferred one (*Welchen der drei Tourguides würden Sie am meisten bevorzugen?* Which of the three tour guides would you prefer most?), 63 participants agree on TG1, mostly because they find that he has a pleasant voice, is easy to understand and competent. Only 14 subjects prefer TG2 because of her pleasant (De.: *angenehm*) and likable (De.: *sympathisch*) voice, and because the voice fits the application context. 10 participants prefer TG3 because the voice is likable (De.: *sympathisch*), natural (De.: *natürlich*) and funny (De.: *lustig*).

Answering the contrasting question which of the tour guides would be the least preferred one (*Welchen der drei Tourguides würden Sie am wenigsten bevorzugen?* Which of the three tour guides would you prefer the least?), 42 participants rate TG3 as least preferable because the voice is difficult to understand (De.: *schwer zu verstehen*), dense (De.: *derb*) and non-professional (De.: *unprofessionell*). For 35 subjects TG2 is the least preferable voice for the application as it is unpleasant (De.: *unangenehm*), difficult to understand, and sounds arrogant. Only 12 subjects consider TG1 as least appropriate, because the voice is artificial (De.: *zu künstlich, Computerstimme*), and hard to listen to (De.: *anstrengend zuzuhören*). See Table 3 for a summary of the social interpretation and evaluation of the three voices.

The results may also reflect two further issues: On the one hand, the voice of TG1, being incorporated into a commercial text-to-speech system, is better

developed than the voices of TG2 and TG3. On the other hand, local or dialectal varieties in general tend to be less comprehensible than the standard variety. In Soukup's study [29], for instance, 70 out of 213 participants brought up the issue of comprehension in relation to the use of dialectal varieties.

Social interpretation		
TG1	**TG2**	**TG3**
professional speaker (22)	elderly person (62)	is a peasant (10)
academic (19)	retired (33)	likes to go to the pub (20)
lives in the city (26)	lives in the city (16)	lives in the city (29)
(Vienna (9))	(Vienna (9))	(Vienna (29))
Local variety and cultural context		
TG1	**TG2**	**TG3**
most preferred speaker (63)	most preferred speaker (14)	most preferred speaker (10)
pleasant voice	pleasant	likable
easy to understand	likable	natural
competent	voice fits the context	funny
least preferred speaker (12)	least preferred speaker (35)	least preferred speaker (42)
artificial	unpleasant	difficult to understand
hard to listen to	difficult to understand	dense
	arrogant	non-professional

Table 3. Summary of the social interpretation and evaluation of the three tour guides.

The questions regarding the appropriateness of the three tour guides were complemented with questions regarding the appropriateness of the dialectal Viennese for the cultural heritage application (*Wie passend ist das Wienerisch des 3. Sprechers für eine Tour im Prunksaal* How well suited is the Viennese of the 3. speaker for a tour in the State Hall?) and which language variety would be best suited (*Gibt es einen besseren Sprachstil als das Wienerische für diese Aufgabe?* Is there a better suited language variety than Viennese for this task?). 63 participants agree that TG3 is rather inappropriate to very inappropriate. 73 claim that standard Austrian German would be best suited.

Appreciation of Viennese language and people in general: Finally, the participants were asked in two separate questions what they think of Viennese language and people in general (*Wie wirkt das Wienerische auf mich?* How am I affected by Viennese language?; *Wie wirken Wiener auf mich?* How am I affected by Viennese people?). The following answers were given: Regarding Viennese language 31 are positive, 30 negative and 13 see it partially positive and partially negative. Regarding Viennese people 32 see them negative, 15 positive and 19 partially positive and partially negative.

5 Conclusion

In summary, we found similar results to Soukup's study [29] regarding the synthetic voices representing the standard (TG1) and the dialectal (TG3) variants, with the Austrian standard being evaluated as most educated, trustworthy, competent, polite and serious, whereas the voice representing the dialectal variety was evaluated as most natural, emotional, relaxed, open minded, with the highest sense of humour, but also most aggressive. The female voice representing a colloquial variety of Viennese (TG2) is in between the standard and the dialectal variant. It comes close to the standard variant in terms of intelligence, gentleness, politeness, and lack of aggression, and close to the dialectal variant in terms of sense of humour and emotionality. However, a word of caution must be added regarding the interpretation of the proximity of TG2 to TG1 on the one hand and to TG3 on the other hand. Identified similarities and differences may be due to the assessment of language variety, but they might as well reflect social evaluation due to sex. To gain better evidence for the one or the other, at least a male colloquial Viennese synthetic voice would be required to be tested against TG1 and TG3. To fully account for sex-specific aspects, additional synthetic voices are needed, including a female standard Austrian and a female Viennese dialectal voice, as well as a male Viennese colloquial voice, which are not available yet.

The analysis of the open questions reveals a clear distinction between TG1, TG2 and TG3. While for TG1, characteristics such as living in the city, professional speaker and academic are perceived as most prominent, it is age for TG2 and regional attribution for TG3. This provides evidence that both language variety and other specific vocal characteristics (such as those referring to age) are relevant for social interpretation and categorization. There may also be other vocal characteristics that influence social interpretation. TG3, for instance, is believed to 'like to go to the pub'. Additionally our results show that differences in the evaluation by male and female participants may be negligible with respect to the present data.

An important lesson from the present study is that, similar to natural voices, language variety together with vocal characteristics of synthetic voices elicit social interpretation and evaluation. This influences the attitudes of human listeners towards artificial speakers in specific ways. Therefore the selection of voice is crucial for communicative agents, and must fit the character's appearance and its behaviour as well as the application context the character appears in.

Acknowledgements

In part, this work has been funded by the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT) under the research programme "FEMtech women in research and technology" within the project "Companions für Userinnen" (C4U).

The authors wish to thank the anonymous reviewers for their valuable and insightful comments.

References

1. Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N.J. Salkind, D.M., Dougherty, B. Frey (Eds.): Encyclopedia of Research Design. Thousand Oaks (CA): Sage, 573–577 (2010);
URL `http://www.utdallas.edu/ herve/abdi-Holm2010-pretty.pdf` (last retrieved 24.6.2012)
2. Cassell, J.: Culture as Social Practice: Being Enculturated in Human-Computer Interaction. In C. Stephanidis (Ed.) Proceedings of HCII, (published as Universal Access in HCI, Part III. Berlin Heidelberg: Springer-Verlag), 303–313 (2009)
3. Clark, R., Richmond, K., King, S.: Multisyn voices from ARCTIC data for the Blizzard challenge. Proceedings of Interspeech, 101–104 (2007)
4. Crowell, C., Scheutz, M., Schermerhorn, P., Villano, M.: Gendered voice and robot entities: perceptions and reactions of male and female subjects. Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, Missouri (2009)
5. Dubinsky, A.J., Skinner, S.J., Whittler, T.E.: Evaluating sales personnel: An attribution theory perspective. Journal of Selling and Sales Management 9 (2), 9–21 (1989)
6. Garrett, P., Coupland, N., Williams, A.: Investigating Language Attitudes. Social Meanings of Dialect, Ethnicity and Performance. Cardiff: University of Wales Press (2003)
7. Heider, F.: The Psychology of Interpersonal Relations. Wiley, New York (1958)
8. Holm, S.: A simple sequential rejective multiple test procedure. Scandinavian Journal of Statistics 6:65-70 (1979)
9. Iacobelli, F., Cassell, J.: Ethnic Identity and Engagement in Embodied Conversational Agents. Proceedings of Intelligent Virtual Agents (IVA), Sept. 17-19, Paris, France, 57–63 (2007)
10. Kelley, H.H.: Causal Schemata and the Attribution Process. General Learning Press, New York (1972)
11. Krenn, B., Sieber, G., Petschar, H.: Metadata Generation for Cultural Heritage: Creative Histories - The Josefsplatz Experience. Proceedings of EVA (Electronic Information, the Visual Arts and Beyond) 2006, Vienna, Austria , 27–34 (2006)
12. Lambert, W.: A Social Psychology of Bilingualism. Journal of Social Issues. 23(2), 91–109 (1967)
13. Lambert, W., Hodgson, R., Gardner, R., Fillenbaum, S.: Evaluational reactions to spoken languages. Journal of Abnormal and Social Psychology. 60(1), 44–51 (1960)
14. Moon, Y., Nass, C.: How 'real' are computer personalities? Psychological responses to personality types in human-computer interaction. Communication Research 23 (6), 651–674 (1996)
15. Moosmüller, S.: Dialekt ist nicht gleich Dialekt. Spracheinschätzung in Wien. Wiener Linguistische Gazette 40-41, 55–80 (1988)
16. Moosmüller, S.: Hochsprache und Dialekt in Österreich. Soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien, Graz, Salzburg und Innsbruck. Wien, Köln, Weimar: Böhlau (1991)
17. Nass, C., Brave, S.: Wired for Speech. MIT Press, Cambridge, MA (2005)
18. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. Journal of Social Issues 56 (1), 81–103 (2000)
19. Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, D.C.: Can computer personalities be human personalities? International Journal of Human Computer Studies 43, 223–239 (1995)

20. Nass, C., Moon, Y., Green, T.: Are computers gender neutral? Gender stereotypic responses to computers. Journal of Applied Social Psychology 27 (10), 864–876 (1997)
21. Nomura, T., Kanda, T., Suzuki, T.: Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. AI & Society, 20(2), 138–150 (2006)
22. Pucher, M., Neubarth, F., Strom, V., Moosmueller, S., Hofer, G., Kranzler, C., Schuchmann, G., Schabus, D.: Resources for speech synthesis of Viennese varieties. Proceedings of LREC, 2010, Malta, 105–108 (2010)
23. Pucher M., Schabus D., Junichi Y., Neubarth F., Strom V.: Modeling and interpolation of Austrian German and Viennese dialect in HMM-based speech synthesis. Speech Communication, 52(2), 164–179 (2010)
24. Rakic, T., Steffens, M.C., Mummendey, A.: Blinded by the accent! The minor role of looks in ethnic categorization. Journal of Personality and Social Psychology 100(1), 16–29 (2011)
25. Ryan, E., Giles, H. (eds.): Attitudes towards Language Variation. London: Edward Arnold (1982)
26. Schermerhorn, P., Scheutz, M., Crowell, C.: Robot social presence and gender: Do females view robots differently than males? In Proceedings of the Third ACM IEEE International Conference on Human-Robot Interaction, Amsterdam, NL, 263–270 (2008)
27. Sormann, M., Reitinger, B., Bauer, J., Klaus, A., Karner, K.: Fast and Detailed 3D Reconstruction of Cultural Heritage. International Workshop on Vision Techniques applied to the Rehabilition of City Centres 2004, Lisbon, Portugal (2004) CD proceedings
28. Soukup, B.: 'Y'all come back now, y'hear!?' Language attitudes in the United States towards Southern American English. VIEWS (Vienna English Working Papers). 10(2), 56–68 (2001)
29. Soukup, B.: Dialect use as interaction strategy: A sociolinguistic study of contextualization, speech perception, and language attitudes in Austria. Wien: Braumüller (2009)
30. Weiner, B.: Motivationspsychologie. Weinheim: Beltz (1994)
31. Quintanar, L., Crowell, C., Moskal, P.: The interactive computer as a social stimulus in human-computer interactions. In Salvendy, G., Sauter, S., Hurrell, J. (eds.), Social ergonomic and stress aspects of work with computers. Elsevier, Amsterdam, 303–310 (1987)
32. Quintanar, L., Crowell, C., Pryor, J., Adamopoulos, J.: Human-computer interaction: A preliminary social-psychological analysis. Behavior Research Methods and Instrumentation 14, 210–220 (1982)
33. Zahn, C., Hopper, R.: Measuring Language Attitudes: The Speech Evaluation Instrument. Journal of Language and Social Psychology 4(2), 113–123 (1985)