

Effects of Language Variety on Personality Perception in Embodied Conversational Agents

Brigitte Krenn¹, Birgit Endrass², Felix Kistler², and Elisabeth André²

¹The Austrian Research Institute for Artificial Intelligence,
Freyung 6/6, A-1010 Vienna, Austria
brigitte.krenn@ofai.at

²University of Augsburg,
Universitätsstr. 2, 86159 Augsburg, Germany
{endrass,kistler,andre}@hcm-lab.de

Abstract. In this paper, we investigate the effects of language variety in combination with bodily behaviour on the perceived personality of a virtual agent. In particular, we explore changes on the extroversion-introversion dimension of personality. An online perception study was conducted featuring a virtual character with different levels of expressive body behaviour and different synthetic voices representing German and Austrian language varieties. Clear evidence was found that synthesized language variety, and gestural expressivity influence the human perception of an agent's extroversion. Whereby Viennese and Austrian standard language are perceived as more extrovert than it is the case for the German standard.

1 Introduction

In the present contribution, we address cultural implications of multimodal expressive behaviours in artificial agents. In particular, we investigate effects of language variety in combination with linguistic and gestural expressivity on the assessment of a virtual agent's personality on the dimension extroversion-introversion. As regards language variety, we concentrate on German and Austrian standard and Viennese dialect. While the combination of linguistic and bodily expression of extroversion in virtual agents has already been assessed in previous work, e.g. (Neff et al. 2010, Isbister and Nass 2000), studying the effects of language variety on an agent's perceived personality is novel.

In the past, synthetic voices available for text-to-speech systems typically have represented standard varieties, e.g. voices for the German, the British English, the American English standard, etc. More recently, localized standard varieties have been made available. Examples are Cereproc's synthetic voices for Scottish, Northern and Southern British, Irish, Catalan, and Austrian German (<https://www.cereproc.com>,

last accessed 7.2.2014). Moreover, the creation of synthetic voices representing varieties of smaller regions is pursued in research contexts. See, for example, work on Austrian varieties (Neubarth et al. 2008, Pucher et al. 2010a, Pucher et al. 2010b). Under these preconditions, localizing virtual agents technically becomes feasible. At the same time, this opens up questions regarding the effects of language variety on the human perception and socio-emotional evaluation of such an agent. Language attitude studies are a well-established means to assess human evaluation of language varieties. Results from a major, recent language attitude study on German varieties (natural voices), for instance, indicate that speakers of Bavarian are perceived as more extrovert (high in spirits) than speakers of the German standard (Gärtig et al. 2010). In a study on synthetic standard Austrian German and Viennese dialectal varieties, (Krenn et al. 2012) demonstrated that: (i) language attitudes towards natural voices transfer to synthesized voices, and (ii) the dialectal Viennese variety is characterized by attributes that also characterize extrovert behaviour. For instance: the dialectal voice compared to the standard Austrian voice is associated with sense of humour, emotionality, self-confidence, open-mindedness and an easy-going nature. These findings together with the fact that Bavarian and Austrian both are varieties of the same German dialect, East Upper German, theoretically back the hypotheses that language variety influences the perceived personality of an agent regarding the dimension extroversion-introversion.

In the next chapter we summarise the theoretical background of our work. Afterwards, we explain how we implemented the introvert and extrovert versions of our embodied conversational agent. In Section 4, we describe our study design and execution, followed by its analysis and interpretation. In the last section, we summarise our findings.

2 Theoretical Background

2.1 Expression of Extroversion-Introversion in Language Variety

Results from Gärtig et al. (2010) indicate that speakers of German standard variety are evaluated as friendly, educated and calm. Speakers of Bavarian are evaluated as even more friendly as speakers of the German standard and in contrast to speakers of the German standard variety Bavarians are perceived as full of spirit. For a summary of the findings see Table 1. 1017 participants rated the "typical German", 501 participants rated the "typical Bavarian".

Krenn et al. (2012) show that language attitudes towards synthesized voices representing local varieties are comparable to language attitudes towards natural voices. 91 subjects of (Austrian) German mother tongue were presented with a semantic differential comprising 19 adjective pairs representing positive and negative extremes on a value dimension, where each pair had to be rated on a 5-point Likert scale. Comparing two synthetic varieties (an Austrian standard male voice and a dialectal Viennese male voice) the following differences could be identified applying a Wilcoxon tests for pairwise comparison and Bonferroni correction for multiple comparisons. Significant differences were found for all dimensions of the semantic differential except for

likeability, friendliness and arrogance. Summing up, the Austrian standard is evaluated as more trustworthy, competent, polite, intelligent, educated and serious as the dialectal Viennese. Whereas the dialectal Viennese is evaluated as more emotional, self-confident, natural, relaxed, with more sense of humour and less strict than the Austrian standard, but also as more aggressive, less gentle and less refined. See for an overview. More information can be found in (Krenn et al. 2012).

Table 1. Evaluation of Standard German speakers and speakers of Bavarian on the dimensions friendliness, educatedness and spiritedness. Summary from Gärtig et al. 2010, cf. pages 103, 106, 109, 113, 116, 119.

Dimension	Typical German N=1017	Typical Bavarian N=501
Friendliness		
(very) friendly	38,7 %	52.5 %
partly	51.4 %	31.7 %
(very) unfriendly	8.3 %	8 %
Educatedness		
(very) educated	45 %	40.8 %
partly	48,9 %	43.4 %
(very) uneducated	4.1	2.6 %
Spiritedness		
(very) spirited	14.5 %	50.6 %
partly	51 %	27.2 %
(very) quiet	30 %	14.1 %

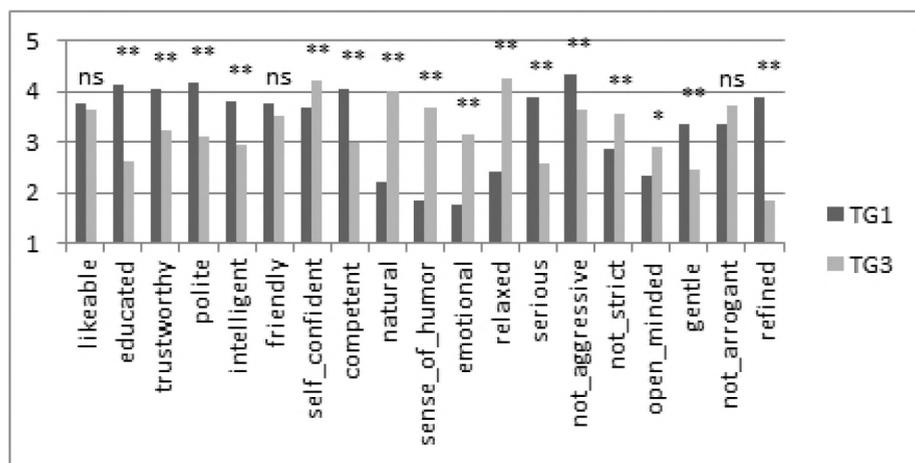


Fig. 1. Pairwise comparison of Austrian standard (TG1) and Viennese dialect (TG3) employing Wilcoxon test and Bonferroni correction; ** $\alpha=0.01$, * $\alpha=0.05$, ns not significant.

2.2 Expression of Extroversion-Introversion in Gestural Correlates

For designing the agents' body behaviours, the following gestural correlates for extroversion/introversion were exploited, see. Indicators for gesture rate, amplitude, gesture direction and body part were taken from (Neff et al. 2010). Similar findings are reported in studies by (Knapp and Hall 2009 and Lippa 1998) stating that extroversion correlates with a higher spatial extent when gesturing. The gestural speed also tends to be higher for extroverts (Lippa 1998 and Brebner 1985) making their gestures look more powerful. As regards self-adaptor gestures, we follow a study on virtual agents from (Neff et al. 2011) where self-adaptors were identified as signalling low emotional stability, as well as earlier findings from (Campbell and Rushton 1978) indicating negative association of self-adaptors and outward-directed gestures, i.e. gestures that signal extroversion.

Table 2. Gestural correlates for extroversion-introversion

	Introversion	Extroversion
Gesture amplitude	Narrow gestures	Wide gestures
Gesture speed	Low	High
Gesture direction	Inward self-contact gestures	Outward gestures
Gesture rate	Low	High
Body part	Elbows/arms close to the body	Elbows away from body
Self-adaptor gestures	Yes	No

3 Building Extrovert-Introvert Agents

For the technical realization, we use the *Virtual Beergarden* running in the AAA application (Damian et al. 2011). In this scenario, an arbitrary number of agents can be loaded that are able to speak and to exhibit animations. For our aim, we employ the *Charlie* character, a male middle-aged virtual agent with a western appearance. The character initially plays an idle animation that includes eye blinking. Other animations are added according to the current body behaviour script.

Verbal behaviour is realized by a text-to-speech component, in which different voices can be used for the characters. The generated speech consists of audio with synchronised lip movements of the characters. To avoid an influence of the semantics of speech on the perceived extroversion of the character, we choose content and wording as linguistically neutral as possible with respect to extroversion/introversion. For our setting, the following text is uttered:

Willkommen im Biergarten. Wir bieten Sitzplätze für hundert Personen und warme oder kalte Speisen. Von allen Plätzen haben Sie einen guten Ausblick auf die Landschaft. Für Firmenfeiern oder private Feste bieten wir Spezialkonditionen an. Und wenn sie kurzfristig mit mehr Personen kommen wollen, macht das auch nichts.

(En gloss: Welcome to the Beergarden. We offer room for 100 persons and cold and hot dishes. From all places, you get a good view on the landscape. For official or private parties we offer special conditions. And in case you wish to bring additional people on short notice, this does not matter.)

3.1 Realization of the Language Varieties

The virtual character is equipped with three different language varieties: standard German, standard Austrian and Viennese dialect. Following (Gärtig et al. 2010 and Krenn et al. 2012), we hypothesize these language variants as representative for different grades of extroversion-introversion, with Viennese dialect being the most extrovert and standard German being the most introvert.

The agents' speech was generated with two different text to speech engines which are:

- The CereVoice SDK (<https://www.cereproc.com/en/products/sdk>) for generating the standard German and standard Austrian utterances. For German, the male voice Alex was used, and for Austrian the male voice Leopold (<https://www.cereproc.com/de/storede>).
- The Festival Multisyn TTS engine (<http://www.cstr.ed.ac.uk/projects/festival/>, Clark et al. 2007) for generating the utterances in Viennese dialect. See http://vsds.ofai.at.vsds_synthesize.cgi, HPO: Festival Unitsel 16 kHz for the respective voice.

3.2 Realization of the Body Behaviour

In the current version of the *Virtual Beergarden* over 70 animations are available for each character. Non-verbal behaviours are divided into gestures, body postures and facial expressions. Predefined gestures and body postures, per se, vary in their expressivity, e.g. an adaptor gesture (such as scratching the nose) has a lower spatial extent compared to a waving movement as used for greeting. In addition to this gesture-inherent expressivity, animations can be further customized to show different levels of expressivity, e.g. an animation can be played with a higher frame rate to increase the speed of the gesture. In previous work, Damian et al. (2011) conducted evaluation studies to test, amongst others, the effects of variations in non-verbal behaviour on the perceived personality (introvert vs. extrovert) of virtual characters in a conversational setting. The same animation technology is employed as in the *Beergarden*. We take this as further evidence for the validity of the animations in the present study to adequately transport extrovert, introvert and neutral behaviour.

Three different non-verbal variations were created to match the text spoken by the agent: introvert, neutral and extrovert. The introvert version is characterised by a lower animation rate compared to the neutral and extrovert version. It includes adaptor gestures and has low spatial extent in gestures and body postures. The neutral version contains more animations than the introvert version, but fewer gestures than the extrovert version. Animations are at a middle level of spatial extent and speed. The extrovert version shows the highest gesture rate, spatial extent, and speed. No adaptor gestures are used in the neutral and extrovert settings. Fig. 2 shows screenshots of the agent exhibiting animations in the extroversion and introversion setting.

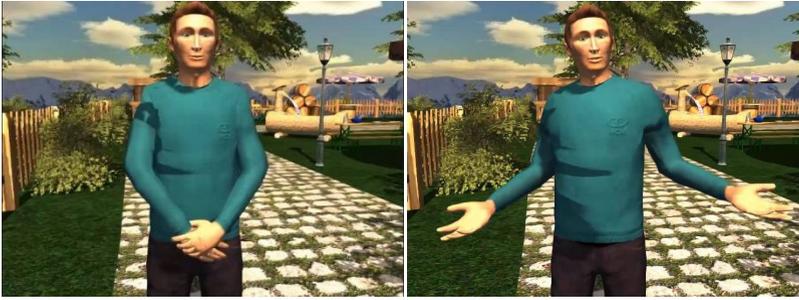


Fig. 2. The Charlie character exhibiting animations in the extrovert (left) and introvert (right) versions

4 The Study

4.1 Experimental Design

We follow Neff et al. (2010) in the experimental design of the study and in the design of extrovert-introvert body behaviour, and extend it by employing different synthetic voices which stand for different language varieties, namely standard German, standard Austrian German and Viennese dialect. The main goal of the present study is to assess how far the perceived extroversion-introversion of a language variety influences the human perception of an agent's extrovert or introvert personality.

Parameter Combinations. We follow a 3x3 design where the synthetic voices/language varieties and bodily behaviours are varied while the agent's text, graphical appearance and setting of the scene are kept constant. This results in 9 settings represented by 9 videos as summarized in Table 3. For the sake of simplicity, from now on we will refer to the number of each video instead of the respective parameters combinations.

Table 3. Combination of parameters as used in the study

Variation in speech/ Non-verbal behaviour	Standard German	Standard Austrian	Viennese Dialect
Introvert	Video 1	Video 4	Video 7
Neutral	Video 2	Video 5	Video 8
Extrovert	Video 3	Video 6	Video 9

Questionnaire. The assertions presented in are adapted from the Big-Five questionnaire published in (Satow 2012), in order to assess the agents' extroversion-introversion dimension as perceived by humans. Whereas the original assertions are made from the human agent's perspective, we have adapted the assertions to be uttered from the observer's perspective in the form of *I believe that the agent in the video* **ASSERTION**. Each assertion is rated on a 7-point Likert scale ranging from 'not at all' (*Trifft überhaupt nicht zu*) to 'fully agree' (*Trifft voll und ganz zu*). See Table 4 for the respective German wording.

In addition, the participants were asked for their age, gender, mother tongue, nationality, where they have mainly lived for the last 5 to 10 years, and whether they had experience with virtual agents and speech synthesis before.

Hypotheses. The videos shown to the participants contain variations in non-verbal behaviour and language variety. The hypotheses are theoretically grounded in earlier findings: (i) East Upper German is perceived by German speakers as extrovert as opposed to Standard German which is perceived as introvert (Gärtig et al. 2010); (ii) Viennese dialect is evaluated as more extrovert as Austrian standard (Krenn et al. 2012); and (iii) gestural expressivity increases the perception of extroversion (Neff et al. 2010).

The hypotheses tested in the present study are:

- **H1:** Language variety increases the perceived extroversion of an agent.
- **H2:** Increased extroversion in non-verbal behaviour leads to an increased perceived extroversion of the agent.
- **H3:** A combination of parameters that are perceived as more extrovert leads to even higher perceived extroversion of the agent.
- **H4:** Viennese dialect is perceived as more extrovert than Austrian and German standard.
- **H5:** German standard is perceived as least extrovert in comparison to Austrian standard and Viennese dialect.

Table 4. Assertions to assess the perceived extroversion-introversion of an agent.

Ich glaube, dass der Agent im Video	gerne mit anderen Menschen zusammen ist. (‘likes to be in company’) ein Einzelgänger ist. (,is a loner‘) in vielen Vereinen aktiv ist. (,is active in many clubs,) ein gesprächiger und kommunikativer Mensch ist. (‘is a talkative and communicative person’) im Grunde oft lieber für sich allein ist. (,prefers to be on his own‘) sehr kontaktfreudig ist. (,is very sociable,) schnell gute Stimmung verbreiten kann. (,can quickly spread a good mood‘) gerne auf Partys geht. (,likes to go to parties‘) unternehmungslustig ist. (is adventurous) gerne im Mittelpunkt steht. (,likes to be in the center of attention‘)
-------------------------------------	---

Table 5. Likert scale rating for the extroversion-introversion assertions as presented to the participants

Trifft überhaupt nicht zu	Trifft größtenteils nicht zu	Trifft eher nicht zu	Weder zutreffend noch unzutreffend	Trifft eher zu	Trifft größtenteils zu	Trifft voll und ganz zu
Applies not at all	Applies not for the most parts	Applies rather not	Neither applies or not applies	Rather applies	Applies for the most parts	Fully applies
1	2	3	4	5	6	7

4.2 Participants

The videos and questionnaires were embedded in a website. Thus, participants were not distracted by a lab setting, but able to watch the videos where and when they liked. In total, 45 people (22 female, 23 male) participated in our study. 22 of the participants have German nationality and 23 have Austrian nationality. Participants were in an age range from 23 to 54 years (\bar{M} 32.2). On a 4-point scale ranging from no experience (1) to very much experience (4), participants reported their previous experience with virtual characters (\bar{M} 1.69, Median 1) and speech synthesis (\bar{M} 1.62, Median 1). All participants are of German mother tongue, except for one whose mother tongue is English and another one who is bilingual (German/Bosnian). Both participants are of Austrian nationality and have lived in Austria for the last 5-10 years, and therefore were included in the analysis.

5 Analysis and Interpretation

A two-way repeated measures ANOVA (using SPSS) with the factors language variety and body behaviour was computed on the participants' evaluation of the perceived extroversion of the agent in videos 1 to 9. As a post-hoc test, Bonferroni corrected t-tests were applied.

The ANOVA showed highly significant effects ($\alpha < 0.01$) of language variety ($F(2, 88) = 30.81$) as well as of body behaviour ($F(1.42, 62.34) = 13.57$ with Greenhouse-Geisser corrections) on the perceived extroversion of the agent. The results confirm H1 and H2. However, there are no significances for joint effects of language variety and bodily behaviour. Accordingly, H3 has to be rejected.

As regards language variety, the pairwise comparisons showed differences ($\alpha < 0.01$) between standard German on the one hand, and standard Austrian and Viennese dialect on the other hand, whereby both standard Austrian and Viennese dialect were evaluated as more extrovert than standard German. Thus H5 is confirmed. H4 must be rejected, as there are no differences between standard Austrian and Viennese.

Although the ANOVA suggested no significances for the interaction between language variety and body behaviour, a closer look at the within-subjects contrasts

revealed a significantly higher difference ($\alpha < 0.05$) in the perceived extroversion of the agent between the Austrian standard combined with the extrovert versus the introvert body, than it is the case for the German standard when combined with the extrovert versus the introvert body ($F(1, 44)=7.06$). For illustration see Fig. 3, the distances in the “rated degree of extroversion” between standard German with introvert and extrovert body, and between standard Austrian with introvert and extrovert body.

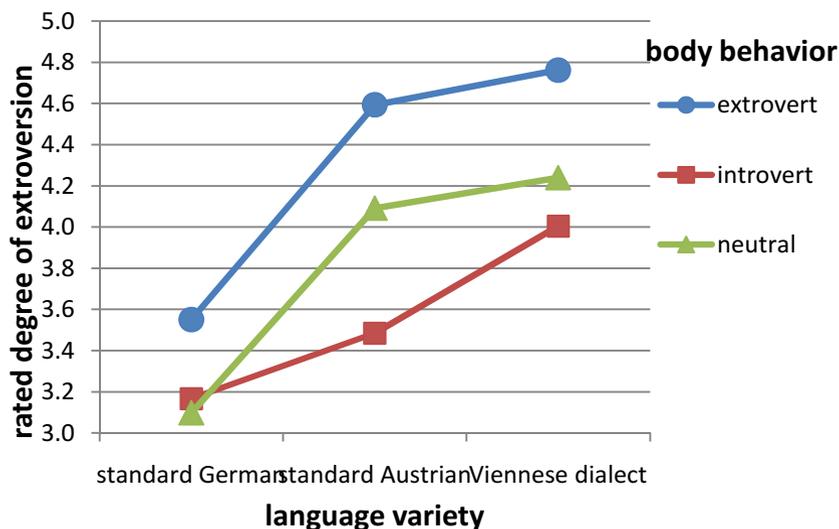


Fig. 3. Rated degree of the agent’s extroversion based on the factors language variety and body behaviour

6 Conclusion

An online-study was conducted with speakers of German and Austrian standard language variety who were to assess a virtual agent’s degree of extroversion based on the agent’s synthesized language variety and its body language. The agent was equipped with synthetic speech representing German and Austrian standard language variety and Viennese dialect, and with gestural behaviour suggesting an extrovert or introvert personality. Clear evidence was found that: (i) Synthesized language variety increases the perceived extroversion of an agent, with Viennese dialect and Austrian standard being perceived as more extrovert than German standard. This confirms findings from language attitude studies on the effects of natural language variety on the perception of a speaker’s personality. Thus the present study provides further evidence for the fact that effects of natural speech transfer to synthesized speech. (ii) Increased extroversion in non-verbal behaviour leads to an increased degree in perceived extroversion of the agent. Whereby aspects such as gesture rate, amplitude and speed, or the presence or absence of self-adaptor gestures influence whether a virtual agent is perceived as extrovert or introvert. Here again, the results from the present study

corroborate findings from earlier works on the effects of non-verbal behaviour on perceived personality. In addition, the data also provide some evidence for joint effects of language variety and non-verbal behaviour, namely the Austrian standard language variety combined with extrovert body behaviour leads to a strong increase of perceived extroversion compared to the effect of Austrian standard combined with introvert body language, whereas the respective effect is significantly smaller when the German standard language variety is combined with extrovert or introvert non-verbal behaviour.

Overall, the results clearly demonstrate the effects language variety, on the one hand, and body language, on the other hand, have on the human perception of an agent's extrovert or introvert personality. Thus the presented work offers respective guidance for the design of artificial agents. Moreover, the current work further supports previous findings (cf. Krenn et al. 2012) that evidence from effects of natural language varieties transfers to synthetic speech. This is relevant for agent designers as their design process may directly profit from existing results of language attitude studies on natural speech. Which may be a factor, given the growing availability of synthetic voices representing language varieties, see for example Cereproc's commercially available synthetic voices for Scottish, Northern and Southern British, Irish, Catalan, or Austrian German.

In future work, we aim to take a closer look at potential differences in the assessment of the agent's personality by the Austrian as opposed to the German user group, and include the gender dimension analysing and comparing the answers from male and female participants separately. In the present study, the assertions referring to the agent's extroversion-introversion in the questionnaire are closely related to human contexts, as they are taken from a typical personality questionnaire which is designed for human self-assessment. See for instance the wording in, in particular the explicit use of the word *Mensch*. This may bias the human ratings towards "neither-nor" answers, especially in combination with the 7-point rating scale applied in the questionnaire. Thus, an analysis of the participants' rating behaviour per assertion is called for in the first place. Moreover, the assertions might need better adaption to an artificial agent context where, for instance, *Mensch* is exchanged by *Charakter* (En.: character). In addition, the rating scale may be changed to 6 or even to 4 in order to (i) prevent the participants from giving "neither-nor" answers as it is possible with the currently used 7-point scale, and to (ii) force answers towards the extroversion or introversion end of the dimension.

References

1. Brebner, J.: Personality theory and movement. Individual differences in movement, pp. 27–41. MTP Press Limited (1985)
2. Clark, R., Richmond, K., King, S.: Multisyn voices from ARCTIC data for the Blizzard challenge. In: Proceedings of Interspeech, pp. 101–104 (2007)
3. Campbell, A., Rushton, J.: Bodily communication and personality. The British Journal of Social and Clinical Psychology 17(1), 31–36 (1978)

4. Damian, I., Endrass, B., Huber, P., Bee, N., André, E.: Individualized Agent Interactions. In: Allbeck, J.M., Faloutsos, P. (eds.) MIG 2011. LNCS, vol. 7060, pp. 15–26. Springer, Heidelberg (2011)
5. Gärtig, A.-K., Plewnia, A., Rothe, A.: Wie Menschen in Deutschland über Sprache denken. Ergebnisse einer bundesweiten Repräsentativerhebung zu aktuellen Spracheinstellungen. amades - Arbeitspapiere und Materialien zur deutschen Sprache. Band 40. Institut für Deutsche Sprache, Mannheim (2010)
6. Gosling, S.D., Rentfrow, P.J., Swann Jr., W.B.: A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality* 37, 504–528 (2003)
7. Knapp, M.L., Hall, J.A.: Nonverbal communication in human interaction. Holt, Rinehart (2009)
8. Krenn, B., Schreitter, S., Neubarth, F., Sieber, G.: Social Evaluation of Artificial Agents by Language Varieties. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 377–389. Springer, Heidelberg (2012)
9. Lippa, R.: The nonverbal display and judgement of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis. *Journal of Research in Personality* (32), 80–107 (1998)
10. Neff, M., Toothman, N., Bowmani, R., Fox Tree, J.E., Walker, M.A.: Don't scratch! self-adaptors reflect emotional stability. In: Vilhjálmsón, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS, vol. 6895, pp. 398–411. Springer, Heidelberg (2011)
11. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS (LNAI), vol. 6356, pp. 222–235. Springer, Heidelberg (2010)
12. Neubarth, F., Pucher, M., Kranzler, C.: Modeling Austrian dialect varieties for TTS. In: Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH 2008), Brisbane, Australia, pp. 1877–1880 (2008)
13. Pucher, M., Neubarth, F., Strom, V., Moosmüller, S., Hofer, G., Kranzler, C., Schuchmann, G., Schabus, D.: Resources for speech synthesis of Viennese varieties. In: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC), Valletta, Malta, pp. 105–108 (2010)
14. Pucher, M., Schabus, D., Yamagishi, J., 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)
15. Satow, L.: Big-Five-Persönlichkeitstest (B5T): Test- und Skaldokumentation (2012), <http://www.drSATOW.de>