Audio-visual Personality Cues for Embodied Agents: An experimental evaluation

E. Krahmer **S. van Buuren**
Communication and Cognition
Faculty of Arts
Tilburg University
The Netherlands
e.j.krahmer@uvt.nl

Zs. Ruttkay
CWI, Centre
for Mathematics
and Computer Science
The Netherlands
zsofia.ruttkay@cwi.nl

W. Wesselink
Dept. of Mathematics
and Computer Science
Eindhoven U. of Technology
The Netherlands
j.w.wesselink@tue.nl

Abstract

We report on an experiment in assigning personality to embodied agents reciting lines from modern poems. Three potential personality cues are investigated: gaze, speech and eyebrow movements. For each cue an introvert and an extrovert variant is defined. Little is known about the perceptual effects of *combining* personality cues. This is addressed by creating personality profiles, consisting of all possible combinations of the three cues of interest. The experimental results show that all three cues have a significant effect on the perception of personality. Concerning combinations of cues we find that including an extravert feature to an otherwise introvert agent does not imply that subjects perceive the agent as more extravert. Rather, we find that an animated character with two arbitrary extravert cues is perceived as significantly more extravert than a character with zero or one extravert cues, but significantly less extravert than one with three extravert cues. Finally, the results show that the personality profile does not influence the quality judgement of the animation.

1 Introduction

If we want embodied agents to be perceived as individuals, one of the first things we should do is give them a recognizable personality. When an embodied agent has a personality, people can process the agents' information better (Laurel 1993). And when the personalities of agent and user match, users tend to be more positive about the interaction (Reeves and Nass 1996, Nass and Lee 2000). So giving embodied agents a personality may have various positive consequences.

Psychological theories of personality are a main source of inspiration for this kind of research. Of the various personality theories, the *trait* theory (originally due to Allport and Odbert 1936 and Allport 1937) seems particularly useful from an embodied agent perspective. Other theories involve biological or surrounding factors which are more difficult to implement in an agent. The trait theory is based on the assumption that the personality structure of a person is primarily based on a number of fixed, characteristic elements (the 'traits'). Currently, the standard personality model in the trait theory consists of 5 personality dimensions (known as **the big five**). These dimensions are Extraversion, Agreeableness, Conscientiousness, Emotional stability and Culture (Tupes and Christal 1961, Norman 1963).[1]

This paper, like most studies of personality for interactive characters, concentrates on the first dimension: extraversion (sometimes also labelled dominance, for instance in Reeves and Nass 1996). It has been claimed that this (together

[1]There appears to be no general consensus on the exact labels of the 5 dimensions. Another common labelling, with a neat acronym, is Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). In this paper, we adopt the nomenclature of Norman.

with friendliness) is one of the two most important of the 5 dimensions (see e.g., McRea and Costa 1989).[2] Obviously, linguistic style is an important cue for personality (Isbister and Nass 2000). But also nonverbal cues have been shown to be relevant. Fukayama et al. (2002), for instance, have shown with an eyes-only interface that certain gaze patterns are typically associated with an introvert personality while others are strong indicators of extraversion. Nass and Lee (2000), to give another example, showed that manipulating pitch and duration parameters in a synthetic speech interface could also create the impression of introversion or extraversion.

So far there has been little research into combinations of nonverbal cues. As a result, it is unclear which cues contribute to what extent and in which combinations to the perception of introversion and extraversion. In addition, it seems likely that there are other facial cues besides speech and gaze which are relevant for the suggestion of personality. It has been shown that expressive gestures and an open body posture are associated with extraversion, while non-expressiveness is associated with introversion (e.g., Isbister and Nass 2000). It seems likely that facial expressiveness, for instance in terms of eyebrow movements, can fulfill a similar role as gestures and body posture.[3] Here we investigate whether facial expressiveness can contribute to the perception of extraversion. In particular, we hypothesize that the presence of eyebrow movements will be associated with extraversion, while the absence of such movements will indicate introversion. We study combinations of all three personality cues —gaze, speech and eyebrows— each with an introvert and an extravert setting. Besides investigating the relative strength of these features and their combinations, we are also interested in the perceived quality of the personalities. Earlier work has shown that people dislike inconsistencies in a character, but it is unclear whether this also applies to combinations of nonverbal cues.

The remainder of this paper is structured as follows. In section 2 we describe the facial animations used in the experiment, and how the various personality profiles were generated. In section 3 the experimental set-up is given. Section 4 describes the results, focussing on cue strength, perception of personality and qualitative evaluation of the personality. The paper ends with a general discussion and some pointers for future research.

2 Materials

For the experiment, a Dutch Talking Head was used. The speech was generated using a Dutch diphone synthesis system. Computer animations were produced with the *CharToon* environment (Ruttkay et al. 1999).[4] A 2D head of a male character formed the basis of the animations. Visual speech is generated on the basis of a set of 48 visemes. Phonemes from the input are matched to corresponding visemes with a sampling rate of 40ms, while intermediate stages are computed using linear interpolation.

Three personality cues were investigated in this experiment: gaze, speech and eyebrow movements. Two gaze patterns were used, one introvert and one extravert (cf. Argyle and Cook 1976, Fukayama et al. 2002). In the extravert gaze pattern (**EG**), the character's eyes are fixed on the user, only interrupted by two blinks to enhance the naturalness of the pattern. In the introvert gaze pattern (**IG**) the character's eyes move to the south-west. This pattern consists of a 300 ms dynamic movement of the eyes, followed by a period of 600 ms in which the eyes are fixed in the lower left corner and finally a 300 ms period in which the eyes move back to their original centered position. In the IG-animation this pattern was repeated twice, interrupted by short period in which the character's eyes are fixed on the user.

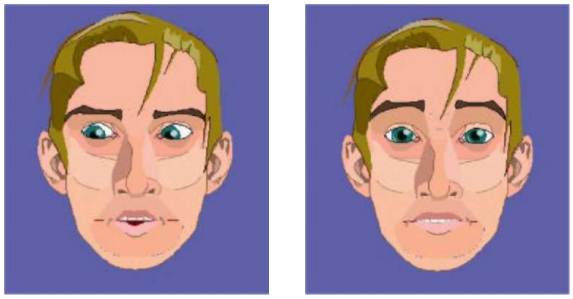
The fundamental frequency (F_0) and duration of the speech were also manipulated in two ways, to create an extravert and an introvert voice (cf. Burgoon and Woodall 1983, Lee and Nass 2000). The extravert speech (**ES**) is characterized by a

[2] Arguably, these are also the easiest to model in an embodied agent. It would probably be harder to develop a cultural or a neurotic agent (but see Colby et al. 1971 for an extreme case). In addition, it seems questionable whether similarity attraction applies for these personalities as well: would a neurotic person prefer a neurotic interface?

[3] For an overview of some of the communicative functions that eyebrow movements can perform, see the seminal Ekman (1979).

[4] See also `http://www.cwi.nl/projects/FASE/`.

Figure 1: Two stills from the Talking Head used in the experiment, illustrating the introvert gaze (left) and the eyebrow movement (right).



broader pitch (F_0) range, more variation in the pitch and a high tempo. The introvert speech (**IS**), on the other hand, has a small pitch range and little pitch variation (it sounds monotonous) and the tempo is 20% slower than that of extravert speech.

Finally, we distinguish introvert brows (**IB**) and extravert brows (**EB**). The former contains no movements of the eyebrows, the latter two eyebrow movements (which coincide with the accented words in the extrovert speech condition). The movements consist of a 100 ms dynamic raising part, a 100 ms static high part and a 100 ms dynamic lowering part. The overall length of the movement is comparable to the average duration of rapid eyebrow movements of human speakers (Cavé *et al.* 1996).

A systematic variation of these three binary cues gives rise to $2^3 = 8$ personality profiles. Each profile was used in one animation. In each animation the talking head uttered a line from a different twentieth-century, 'modern' Dutch poem. The assumption is that these abstract lines give the subjects no clues about the personality of the character. We opted for using different lines for different personality profiles in an attempt to make the experiment less monotonous for the subjects. All lines have a comparable syntactic structure and a similar length. They were randomly assigned to personality profiles. The resulting stimuli are (poets between brackets, English translations by the authors of this paper):[5]

[5]The animations used in the experiment can be viewed at `http://homepages.cwi.nl/~zsofi/stim.html`.

IS+IG+IB Het schip van de wind ligt gereed voor de reis (Marsman). *The ship of the wind is ready for the journey*;

IS+EG+IB Weer heeft de boer zijn beste os gedood (Sontrop). *Once more, the farmer has killed his best ox*;

ES+IG+IB In het onland stond een hert zo groot als god (Achterberg). *On the marshy ground stood a deer as large as god*;

IS+IG+EB Bitter smaakt het kruid der herinnering (Claus). *Bitter tastes the herb of remembrance*;

ES+EG+IB De wijn is drinkbaar dankzij het glas (Mulisch). *The wine is drinkable thanks to the glass*;

ES+IG+EB In de weiden grazen de vreedzame dieren (Marsman). *In the meadows the peaceful animals graze*;

IS+EG+EB Weer gaat de wereld als een meisjeskamer open (Rodenko). *Once more, the world opens like a girls' chamber*;

ES+EG+EB Alles van waarde is weerloos (Lucebert). *Everything of worth is defenceless*.

3 Experimental setup

Subjects of the experiment were 24 native speakers of Dutch, between 18 and 52 years old. 15 subjects were male and 9 female. They were told that the experiment was about assigning 'human-like' properties to animated characters. Naturally, they were not informed about the kinds of cues they could pay attention to. The experiment was individually performed and self-paced. Subjects could watch and listen to each stimulus as often as they desired. Before the actual experiment, subjects entered a brief training session (consisting of two stimuli) to make them acquainted with the materials and the setting of the experiment. No feedback was given on the 'correctness' of their answers and there was no further communication with the conductor of the experiment. The experiment itself consisted of the 8 stimuli described in the previous section. The stimuli were presented

Table 1: Interaction between gaze and brows for mean total personality scores.

	extravert brows	introvert brows
extravert gaze	23.28	17.28
introvert gaze	17.65	14.34

in two different random orders, to compensate for possible learning effect. On average, subjects required 20 minutes to finish the experiment.

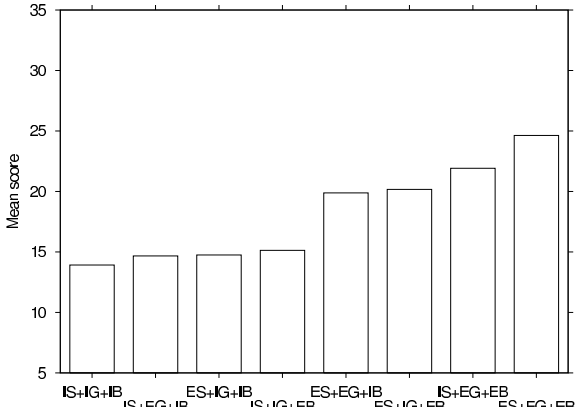
For each stimulus, subjects were asked to fill in a short two-part questionnaire. In the first part, they had to rate the perceived personality of the animations. This was done using a 7-point semantic differential with 5 pairs of Wiggins' (1979) adjectives for introversion and extraversion at the extremes (shy/enthousiastic, inward/outgoing, bashful/perky, unrevealing/open, non-vivacious/vivacious). In the questionnaire, the adjectives were mixed. During processing of the results, introvert adjectives were mapped to 1 and extrovert ones were mapped to 7. The total personality score that a subject assigned to an animation was computed by summing over the scores for the 5 questions and thus ranges from 5 (completely introvert) to 35 (extravagantly extravert). The second part of the questionnaire consisted of 3 questions about the realization of the personality. Subjects had to score the quality, understandability and naturalness of the character. The quality score for a subject is computed per stimulus by summing over the scores for the 3 individual questions and ranges from 3 (very poor) to 21 (very good). All subjects rated all animations.

4 Results

4.1 Cue influence

A multiple analysis of variance (MANOVA) with repeated measures was carried out to determine which cues had an influence on the perception of personality. As it turns out, all three cues have a highly significant effect on the perceived personality. For speech: $F(1,23) = 17.93, MSE = 31.83, p < .001$. For gaze: $F(1,23) = 31.12, MSE = 28.28, p < .001$. And for brows: $F(1,23) = 48.26, MSE = 21.56, p < .001$.

Figure 2: Mean total personality scores for the eight animations ($n = 24$). The personality scale ranges from 5 (very introvert) to 35 (very extravert).



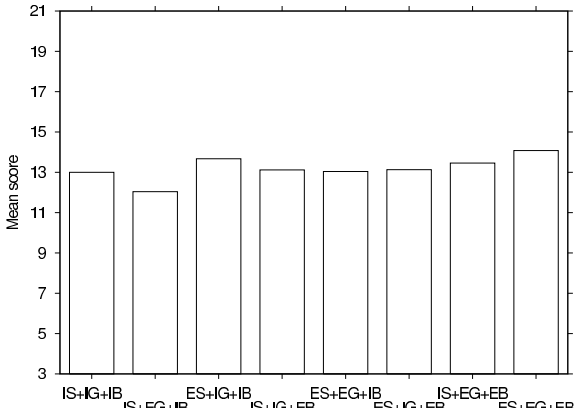
Two interactions were found. One between gaze and brows ($F(1,23) = 5.27, MSE = 16.43, p < .05$) and one between gaze, brows and speech ($F(1,23) = 11.13, MSE = 12.14, p < .01$). This suggests that gaze and brows reinforce each other. As can be seen in Table 1, an animation which contains extravert brows is perceived as more extravert (i.e., scores higher on the personality score) when it is combined with extravert gaze, and similarly, an animation with introvert brows is perceived as more introvert when it is combined with introvert gaze.

4.2 Personality perception

Figure 2 contains the mean total scores of the 5 personality questions. The consistency of these 5 items was high (Cronbach's α = .82). As one would expect, the animation containing only introvert cues (IS+IG+IB) scores lowest (most introvert) on the personality scale, while the animation containing only extravert cues (ES+EG+EB) scores highest (most extravert) on the scale.

Interestingly, the set of eight animations can be partitioned into three groups. The first group contains all the animations with at most one extravert cue (i.e., IS+IG+IB, IS+IG+EB, IS+EG+IB and ES+IG+IB). None of these four animations scores significantly different from one of the others on a two-tailed t-test; the biggest difference

Figure 3: Mean total quality scores for the eight animations ($n = 24$). The quality scale ranges from 3 (very bad) to 24 (very good).



is that between IS+IG+IB and IS+IG+EB (an average difference of 1.21), and is not significant ($t(1, 23) = 1.17, p = .25$).

The second group contains all the animations with exactly two extravert cues (i.e., ES+IG+EB, ES+EG+IB and IS+EG+EB). All three are perceived as significantly more extravert than the animations from the first group. The smallest difference is between IS+IG+EB and ES+EG+IB. The latter scores 4.75 points higher than the former, and this difference is statistically significant ($t(1, 23) = 3.23, p < .01$). But the scores for the three animations in the middle group do not differ significantly from each other. The biggest difference is that between IS+EG+EB and ES+EG+IB; the former is on average 2.04 points higher on the extraversion scale, but this difference is not significant ($t(1, 23) = 1.15, p = .26$).

The third group, finally, contains the animations with three extravert cues (i.e., only ES+EG+EB). This animation is perceived as 2.71 points higher on the extraversion dimension than the most extravert of the second group (IS+EG+EB), which is a significant difference ($t(1, 23) = 2.1, p < .05$).

4.3 Quality of personality

Subjects also had to rate the quality of the way the personality was modelled in the character. Three questions were answered on a 7-point semantic differential scale. Again, the internal consistency for this scale is high (Cronbach's α = .91). The results are given in Figure 3. Interestingly, the personality profile has no influence at all on the quality assessment. Neither speech ($F(1, 23) = 2.34, MSE = 6.74, p = .14$), nor gaze ($F(1, 23) = .04, MSE = 6.64, p = .85$), nor brows ($F(1, 23) = 1.78, MSE = 7.02, p = .20$), nor any combination has a significant effect on the quality judgement.

5 Discussion and conclusion

In this paper we have studied all combinations of three audio-visual cues (gaze, speech and brows) that may be used to suggest a personality for an embodied facial agent. For each of these cues, an introvert and an extravert version was defined, based on earlier research and claims in the literature. The experiment showed that each of these cues has a highly significant influence on the perceived personality of the agent, where the effects of gaze and brows seem to reinforce each other. This might be due to the fact that the eyes and brows are located in the same facial region.

That speech has an influence confirms the results of the speech-only experiment of Nass and Lee (2000); that gaze has an influence as well confirms the eyes-only experiment of Fukayama et al. (2002). That eyebrows may also contribute to the perception of personality is perhaps more surprising. It seems likely that this is a particular manifestation of the general influence of *expressiveness*; a more expressive face is probably associated with extraversion. Interestingly, the experiment showed that an animation with only introvert cues is not necessarily perceived as more introvert than one which includes extravert gaze, speech or brows. An animation with two extravert characteristics is perceived as more extravert than one with a single extravert feature, and an animation with three extravert features is perceived as more extravert still.

The evaluation showed that the personality of the character had no influence on the quality assessment. In itself that is a good thing; it means that one can select the desired personality for an embodied agent, without risking poorer evaluation results. Note that this finding seems to contradict the observation from Reeves and Nass (1996) that

people prefer an interface with a personality similar to their own. Even though we did not test the personality of our subjects, we would expect some of them to be relatively extravert and others relatively introvert. So, it could be the case that the more introvert of our subjects tend to be more positive about the fully introvert animation (IS+IG+IB), while the more extravert of our subjects would prefer the fully extravert animation (ES+EG+EB). This would result in more variation between the subjects (and thus in high standard deviations). However, a post hoc analysis of the results showed that this is not the case. This suggests that the personality of subjects does not play a large role in the quality evaluation here. Perhaps this is caused by the relative unnaturalness (modern poetry) of the current experiment, and maybe also by the fact that here, as opposed to the studies of Reeves, Nass and co-workers, there is no real interaction between subject and agent.

The evaluation result also seems to run counter to earlier observations that people disprefer inconsistent embodied agents. Isbister and Nass (2000), for instance, found that people were negative about an embodied agent when the content and form displayed different personalities (i.e., an extravert linguistic style and an introvert posture). Arguably, some of our personality profiles are also 'inconsistent' (e.g., because the speech is extravert and the gaze introvert). We conjecture that inconsistencies between various nonverbal cues (as we have studied here) are less 'offensive' than inconsistencies between verbal and nonverbal cues.

Finally, we would like to raise two issues that we hope to address in future research. The first is concerned with the question: what is needed to obtain a *very* extravert score for an embodied agent? We have seen that even the profile containing only extravert cues (ES+EG+EB) is still some 10 points below the maximum score on the extraversion scale. It might be that the general quality of the animations used in this experiment accounts for this (there could be a better fit between the synthetic voice and the computer graphics, for instance) or the fact that we concentrate on the upper half of the face (eyes, brows) and do not model mouth expressions (smiles) or global head movements. But, although we intend to work on these aspects in the future, we believe that for higher extraversion scores we really need to combine audio-visual cues with verbal behavior (linguistic register) and interactive behavior (e.g., an extravert agent is more likely to lead in the interaction than an introvert one).

The second topic for future research is the influence of visual appearance on personality perception. Folk wisdom has it that certain kinds of faces are perceived as more introvert or more extravert than others. In the experiment described above we tested the audio-visual personality cues with one facial agent. We are currently redoing the experiment, using the same cues, but various different kinds of faces. A somewhat related question concerns the effect on personality perception of down-scaling the quality of the animation. For mobile applications, for instance, there is only a small display and the frame rate can be rather low. Typically, the speech quality can be better preserved under such conditions. It would be interesting to study the effects of the different cues in such circumstances. We hypothesize that speech cues would be relatively more important than gaze or brows under those adverse conditions, but on the other hand *exaggerating* the visual cues could perhaps redress the balance.

Acknowledgements

Thanks are due to Han Noot for his help in making the animations, to Mariët Theune for her useful comments on a previous version of this paper, and to Antal van den Bosch and Rein Cozijn for their assistance with the statistical analyses.

References

Allport, G. and Odbert, H. (1936), Trait names: a psycho-lexical study, *Psychological Monographs* 47: 211.

Allport, G.(1937), *Personality: A psychological interpretation*. New York: Holt and Company.

Argyle, M. and Cook, M. (1976), *Gaze and Mutual Gaze*, London: Cambridge University Press.

Burgoon, J. and Woodall, W. (1983), Talking fast and changing attitudes: a critique and clarification, *Journal of nonverbal behavior*, 8:126-141.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R., 1996, About the relationship between eyebrow movements and F_0 variations, *Proceedings ICSLP*, Phildelphia, pp. 2175-2179.

Colby, K., Weber, S., Hilf, F. (1971), Artificial Paranoia, *Artificial Intelligence* 2(1): 1-25.

Ekman, P. (1979), About brows: emotional and conversational signals, in *Human ethology*, M. von Cranach et al. (eds.), Cambridge University Pressm pp. 169-202.

Fukayama, A., Hagita, N., Mukawa, N. Ohno, T. and Sawaki, M. (2002), Messages embedded in gaze of interface agents: impression management with agents' gaze, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2002)*, Minneapolis.

Isbister, K. and Nass, C. (2000), Consistency of personality in interactive characters: verbal cues, nonverbal cues and user characteristics, *International Journal of Human Computer Studies* 53:251-267.

Laurel, B. (1993), *Computers as theater*, Reading, Mass: Addison-Wesley.

McCrae, R. and Costa, P. Jr., (1989), Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality, *Journal of Personality* 57(1): 17-40.

Nass, C. and Lee, K. (2000), Does computer-generated speech manifest personality? An experimental test of similarity-attraction, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2000)*, The Hague.

Norman, W.T. (1963), Toward an adequate taxonomy of personality attributes, *Journal of Abnormal and Social Psychology*, 66: 574-583.

Reeves, B. and C. Nass (1996), *The Media Equation: How people treat computers, television, and new media like real people and places*, Stanford: CSLI Publications.

Ruttkay, Zs., ten Hagen, P., Noot, H (1999), CharToon; A system to animate 2D cartoon faces, *Proceedings Eurographics*.

Tupes, E. and Christal, R. (1961), Recurrent Personality Factors Based on Trait Ratings (ASD-TR-61-97). Lackland Air Force Base, TX: Aeronautical Systems Division, Personnel Laboratory.

Wiggins, J. (1979), A psychological taxonomy of trait-descriptive terms, *Journal of Personality and Social Psychology* 37:395-412.