

Tilburg University

Real vs. acted emotional speech

Shahid, S.; Krahmer, E.J.; Swerts, M.G.J.

Published in:

Proceedings of the 4th International Conference on Speech Prosody

Publication date:

2008

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Shahid, S., Krahmer, E. J., & Swerts, M. G. J. (2008). Real vs. acted emotional speech: Comparing south asian- and caucasian speakers and observers. In *Proceedings of the 4th International Conference on Speech Prosody* (pp. 669-772). Unknown Publisher.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Real vs. Acted Emotional Speech: Comparing South-Asian and Caucasian Speakers and Observers

Suleman Shahid, Emiel Krahmer & Marc Swerts

Department of Communication and Information Sciences,
Tilburg University, The Netherlands

{S.Sahid/E.J.Krahmer/M.G.J.Swerts}@uvt.nl

Abstract

Both acted and real emotional audiovisual speech was collected from 50 Caucasian speakers (from The Netherlands) and 45 South-Asian speakers (from Pakistan), using a novel adaptation of the Velten technique, in which some participants are asked to act as if they are in some emotional state, while these emotions are really induced in other participants. Generally, the acted conditions did not lead to systematic non-neutral emotions, while the non-acted conditions lead to the intended emotions being induced in both the Dutch and the Pakistani speakers (where the latter seemed to feel the emotions more strongly). Next we performed a series of perception experiments, in which Dutch and Pakistani observers were asked to judge the emotional state of Dutch and Pakistani speakers. It was found that acted emotions of speakers from both cultures were perceived as stronger than their non-acted counterparts by both Dutch and Pakistani observers. Interestingly, for Dutch speakers the negative emotions were perceived as relatively strong, whereas for the Pakistani speakers the positive emotions stood out perceptually.¹

Index Terms: Emotional speech, Cross-cultural Emotion Production and Perception, Acting, Audiovisual Speech

1. Introduction

A speaker's emotional state can be perceived both from auditory cues (in the speech signal) and visual cues (such as facial expressions). It is generally acknowledged that there exist both similarities and differences in cross-cultural emotion recognition (see e.g., Eifembein and Ambady 2002, Eifembein et al. 2002, Mandal 1987). A substantial amount of research focusing on the recognition of emotions in cross-cultural settings use posed facial photographs (Eifembein et al., 2002), and this work has revealed that there exist significant differences among different cultures. In one study by Izard (1971), for instance, it was shown that the percentage of correctly identified facial photographs was highest among American and European participants whereas it was lower in the Japanese group and the lowest in the African group.

The use of posed facial photographs for emotion studies has a very long research tradition, dating back to the influential work of Darwin (1872) and Ekman (1972), and also in speech research much work is based on "acted" emotions (see e.g., Scherer 2003). It is unclear to what extent acted emotions are representative of real emotions. There is

some suggestive evidence that acted emotions are different from real ones (e.g., Vogt and André 2005) and are perceived as stronger than non-acted ones (e.g., Wilting et al., 2006). Given that cultures differ in the way they feel and express emotions, the question arises to what extent differences between acted and non-acted emotions can be found in different cultures, and to what extent viewers from different cultures perceive such differences between emotions.

The aim of this paper is to look in detail at the difference between acted and real emotions from both a speaker and an observer perspective, comparing South-Asian and Caucasian speakers. For this, we used a novel adaptation of the Velten (1968) technique, first applied to Dutch by Wilting et al., (2006). The Velten technique is used to induce positive or negative emotional states by letting participants read sentences with an increasingly strong emotional content. The Velten technique is widely used, with varying degrees of success, but the method seems very useful for audiovisual speech research.

In this paper we report on two experiments: a new data collection with Pakistani speakers (and comparing it with Dutch speakers collected by Wilting et al. 2006) of both acted and non-acted emotional speech (section 2), and a series of cross-cultural perception experiments in which Dutch and Pakistani observers are asked to rate the perceived emotional state of the speakers from the first experiment (section 3).

2. Experiment I: Data collection

2.1. Method

2.1.1. Design

Experiment 1 had a between participants design with Condition (Positive, Negative, Act Positive, Act Negative and Neutral) and Culture (Dutch, Pakistan) as independent variables and Emotional state as the dependent variable.

2.1.2. Participants

95 speakers participated: 50 Dutch and 45 Pakistani ones evenly and randomly distributed over conditions. Of the 50 Dutch participants, 19 were male and 31 female, with an average age of 27 (range 19-52), and all were students and colleagues from Tilburg University. Of the 45 Pakistani speakers, 28 were male and 17 female, with an average age of 24 (range 19-35), most of them students and staff members from the University of Central Punjab, Lahore. There were no actors among the participants, and none had any involvement in audiovisual speech or emotion research. All participants signed a consent form, stating that their recordings could be used for research purposes.

¹ Thanks are due to Lennard van de Laar for technical assistance, to Emmett Velten for discussion of the eponymous technique and especially to Janneke Wilting for her help in conducting the Dutch data collection experiment.

2.1.3. Materials

To elicit audiovisual expressions of emotions, we used an adaptation of the Velten technique. In the original Velten (1968) version, participants are offered 60 sentences in one of three conditions: Neutral, Positive and Negative. The neutral sentences (e.g., “Mandarin is the official language of China.”) are intended to induce a neutral emotional state. The Positive sentences start neutral as well (“Today is neither better nor worse than any other day”), but gradually get more positive, with “God, I feel great!” as the final sentence. The Negative sentences start with the same neutral sentence, and gradually get more negative with as a final sentence “I want to go to sleep and never wake up”. For our experiments, we made a selection of 40 sentences per condition, where original sentences referring to specifics (e.g., parents, college, religion etc) were removed, and neutral sentences tailored to the American context (e.g., “There is a large rose-growing center near Tyler, Texas”) were replaced with Dutch and Pakistani alternatives. We made sure that the order of sentences in the Positive and Negative conditions was retained. All selected sentences were first literally translated into Dutch and Urdu (the official language of Pakistan), and then revised to make sure the sentences were easy to read and pronounce. For both Dutch and Urdu, the sentences were presented to 3 native speakers who checked their naturalness. For further details on the Dutch experiment see Wilting et al., (2006).

2.1.4. Procedure

The experiment was conducted on the premises of the University of Central Punjab, Lahore, Pakistan and Tilburg University for Pakistani and Dutch speakers respectively. Participants took part in the experiment one at a time, and performed the experiment in a quiet room (in the case of Pakistani speakers a fully air-conditioned room was used so that the participants were not bothered by the summer's high temperature). Participants were asked to sit on a chair placed in front of a desk on which a laptop computer was placed. The laptop was lifted 15 cm above the table surface, so that the screen was roughly positioned at the participants' eye level. Behind the laptop, a digital video camera was positioned in such a way that it could record the participants' full face and upper part of the body.

Besides the three regular conditions (Positive, Negative and Neutral) described by Velten (1968) for the induction of real emotions, two additional conditions were introduced in the experiment. In the first condition, positive sentences were shown to participants but the participant were asked to utter these sentence as if they were in negative mood (Act Negative) and in the second condition, negative sentences were shown to participants but they were asked to utter these sentence as if they were in positive mood (Act Positive).

Before starting the experiment, the experimenter told the participants that the purpose of the experiment was to check the effect of emotion on memory recall. The instructions for the experiment were a slightly adapted version of the original Velten instructions. These instructions were shown on the computer screen and participants were requested to read them first silently and then out aloud to make sure that they practiced the instructions before the experiment started.

Once the participants were ready and the instructions were clear to them, the experimenter left the room and the experiment started. During the actual experiment phase, each sentence was shown on the computer screen for 20 seconds. The participants were asked to first read each sentence silently and then out loud. The camera was continuously recording during this phase. The duration of this phase was 800 seconds (40 sentences x 20 seconds).

Table 1: *Induced emotional state on a 7 point scale (1=very negative, 7=very positive) for Dutch and Pakistan speakers (standard deviations between brackets).*

Condition	Dutch (s.d)	Pakistan (s.d)
Positive	5.65 (0.63)	5.89 (0.62)
Act Positive	4.77 (1.23)	5.26 (0.97)
Neutral	4.95 (0.87)	4.15 (0.47)
Act Negative	4.92 (1.20)	4.24 (1.18)
Negative	3.85 (0.93)	2.91 (0.33)

Once this phase was ended, participants were requested to fill in a short questionnaire assessing their emotional state ("At this moment, I feel ...") due to Mackie and Worth (1989), consisting of six bipolar semantic differential scales. For processing positive adjectives were mapped to 7 (“very positive”) and negative ones to 1 (“very negative”). Once the participants finished this phase, they performed a dummy recall test as recall was supposed to be the purpose of the emotion induction. The results of the memory recall test were not further analysed. At the end of the experiment, the experimenter informed participants about the real purpose of the test. All the Dutch participants were given a candy bar and Pakistani participants a cold juice/ice cream (due to hot weather) as a token of gratitude.

2.2. Results

Figure 1 shows the representative stills of speakers from all four categories (Positive, Act Positive, Negative and Act Negative) and both cultures (Pakistani and Dutch).

The internal consistency of the questionnaire was measured with Cronbach's α and was very good for both the Dutch ($\alpha = .92$) and the Pakistani ($\alpha = .93$) participants. Table 1 gives the induced emotion scores. The first thing to notice for both Dutch and Pakistani participants is that they feel most positive in the Positive condition, and most negative in Negative condition, indicating that the Velten technique worked as intended for both cultures. It is also rather interesting to note that the self-reported scores for the Pakistani speakers are more extreme, in both the positive and negative direction. The acted conditions do not seem to lead to systematic induced emotions, and for these conditions the standard deviations are higher than for the non-acted conditions.

A univariate analysis of variance (ANOVA) confirmed that for the Dutch speakers Condition had a significant effect on the induced emotional state ($F(4, 45) = 4.65, p < .005, \eta^2_p = .288$). A Tukey HSD post hoc test revealed that only the scores for the Positive and the Negative condition differed significantly (at the $p < .001$ level). Also for the Pakistani speakers, condition had a significant effect ($F(4, 40) = 19.274, p < .001, \eta^2_p = .658$). A Tukey HSD post hoc test revealed that most conditions differed significantly, the exceptions being the differences between Act Positive and Positive, between Act Negative and neutral, and between Act Negative and Act Positive.

2.3. Conclusion

The Velten method was found to work very well both for the Dutch and the Pakistani participants; in the non-acted conditions, the intended emotions were indeed induced, which indicates that the translations into Dutch and Urdu and the reduction of sentences from 60 to 40 did not reduce the impact of the approach. The method seemed to work somewhat better for Pakistani speakers, whose self-reported



Figure 1: Representative stills for (L-R): Positive, Act Positive, Negative and Act Negative, for Dutch (top) and Pakistani (bottom) speakers

emotion scores were more extreme than those of the Dutch speakers. Interestingly, the acted emotions did not induce a systematic emotional state in participants in either culture. In the next section, we will report on a series of perception experiments both with Dutch and Pakistan observers.

3. Experiment II: Perception

3.1. Method

3.1.1. Design

The second experiment had mixed design with Condition (Positive, Negative, Act Positive, Act Negative and Neutral) as within-participants variable and Experiment (Dutch viewers judging Dutch speakers, Dutch viewers judging Pakistani speakers, Pakistani viewers judging Dutch speakers and Pakistani viewers judging Pakistani speakers) as between-participants variable and Perceived Emotional State as the dependent variable.

3.1.2. Participants

135 people participated (all different from the speakers in the first experiment), 65 Dutch viewers and 70 Pakistani viewers, with a roughly equal number of men and women in both cultures.

3.1.3. Materials

For each of the 50 Dutch and 45 Pakistani speakers recorded in the first experiment the last sentence was selected. These sentences are assumed to be most representative of the induced emotions as they were the final, most emotion-laden sentences and were uttered just prior to filling in the questionnaire. The sentences were presented to viewers without sound, to prevent them from using lexical cues for their judgement.

3.1.4. Procedure

Four group experiments were conducted: Dutch viewers judging Dutch speakers, Dutch viewers judging Pakistani speakers, Pakistani viewers judging Dutch speakers and Pakistani viewers judging Pakistani speakers, with essentially the same procedure in each case.

For each of the four experiments, groups of participants were invited into a quiet classroom where the computer screen was projected on the front white wall using a beamer. Participants were told, depending on the experimental condition, that they would see 50 Dutch or 45 Pakistani speakers in different emotional states and that their task would be to rate the perceived emotional state of the speakers on a 7 point valency scale where 1 is “very negative” and 7 is “very positive” and 4 is “neutral”. Participants were not informed about the fact that twenty of the speakers were acting in the video. Each stimulus was preceded by a number which indicated which stimulus was coming up, and followed by a six second gap during which a white screen was displayed and which participants could use to fill in their scores on the answer form. Before starting the real experiment, a short training session was shown in which 5 speakers were presented uttering one of the prefinal sentence (one for each condition) to make participants familiar with the stimuli and the task at hand. If everything was clear, then the actual experiment started which lasted for approximately 10-12 minutes depending on the experimental condition. During the experiment there was no interaction between participants and the experimenter.

3.1.5. Statistical analysis

The data were analysed with a mixed within- and between-participants analysis of variance (ANOVA) with 2 factors (Condition and Experiment). Mauchly's test for sphericity was used to test for homogeneity of variance, and if significant we applied a Greenhouse-Geisser correction on the degrees of freedom (for the purpose of readability we report the normal degrees of freedom in these cases). The Bonferroni correction was applied for multiple pairwise comparisons.

Table 2: Average overall perceived valency scores (1 = very negative, 7 = very positive) as a function of the five conditions, with std. errors between brackets and with 95% Confidence Intervals.

Condition	Mean (S.E)	95% CI
Positive	4.491 (0.033)	(4.426, 4.557)
Act Positive	4.961 (0.035)	(4.892, 5.029)
Neutral	3.243 (0.043)	(3.158, 3.328)
Act Negative	2.850 (0.045)	(2.762, 2.939)
Negative	2.998 (0.040)	(2.919, 3.078)

3.2. Results

The overall results are given in Table 2. A significant main effect of Condition was found ($F(4, 524) = 936.035, p < .001, \eta^2_p = .877$, after a Greenhouse-Geisser correction). All pairwise comparisons between conditions were statistically significant at the $p < .001$ level, with the acted conditions receiving the most extreme scores. No significant main effect of Experiment was found ($F(3, 131) = 1.334, n.s.$), as one would expect. Interestingly, however, we did find a significant interaction between Condition and Experiment ($F(12, 524) = 25.257, p < .001, \eta^2_p = .366$). This interaction can be explained via inspection of Table 3. First consider Dutch viewers judging Dutch speakers: the most extreme scores are obtained for the acted conditions, where especially the difference between Act Negative and Negative is clear (the difference between Act Positive and Positive is much smaller). Essentially the same pattern can be seen for the Pakistani viewers watching the same Dutch speakers. Next consider the Pakistani viewers watching Pakistani speakers; here we also see an effect of acting, but on the other end of the scale; here it is Act Positive that is perceived much stronger than its non-acted counterpart. Act Negative is perceived as more positive than the Negative and Neutral conditions; the latter two conditions are perceived as fairly negative (perhaps because Act Positive receives relatively high scores). When the Dutch viewers watch the Pakistani speakers the pattern is essentially the same; they also perceive Act Positive as by far the most positive condition. Compared to the Pakistani viewers they are less good at classifying the non-acted Positive sentences ($M = 3.8$ for Dutch viewers and $M = 4.6$ for Pakistani viewers). They seem to make no distinction between the remaining three conditions (Neutral, Act Negative and Negative).

Table 3: Average perceived valency scores as a function of Condition and Experiment.

Speaker	Condition	Dutch Viewers (sd)	Pakistani Viewers (sd)
Dutch	Positive	4.810 (0.36)	4.723 (0.41)
	Act Positive	4.860 (0.40)	4.866 (0.44)
	Neutral	3.523 (0.43)	3.416 (0.50)
	Act Negative	2.537 (0.49)	2.585 (0.62)
	Negative	3.060 (0.41)	3.065 (0.54)
Pakistani	Positive	3.800 (0.36)	4.633 (0.35)
	Act Positive	4.991 (0.29)	5.126 (0.39)
	Neutral	3.107 (0.50)	2.926 (0.53)
	Act Negative	3.071 (0.41)	3.207 (0.41)
	Negative	3.058 (0.46)	2.811 (0.38)

3.3. Conclusion

The series of cross-cultural perception experiments gave very similar results for Dutch and Pakistani observers, although the latter viewers, judging Dutch speakers, are a somewhat more accurate than the Dutch viewers judging Pakistani speakers. Overall, acted emotions are perceived as more extreme than their non-acted counterparts, with the Neutral condition perceived precisely in between. Interestingly, the effects of acting were especially strong in the negative condition for the Dutch speakers, and in the positive condition for the Pakistani speakers.

4. General Discussion

We have described two series of experiments on the production and perception of acted and real emotions, in a cross-cultural setting comparing Caucasian (Dutch) and South-Asian (Pakistani) speakers and viewers. It was found, for both cultures, that acted emotions are not truly felt (Experiment I), but are perceived stronger than their non-acted counterparts (Experiment II); this pattern is even clearer for Pakistani speakers than for Dutch ones (cf. Wilting et al., 2006). Data revealed that negative acting stands out for Dutch participants and positive acting for Pakistani participants. Interestingly, this same bias (negative for Dutch, positive for Pakistani) was found in a completely different experimental setting (game playing) and for a completely different group of participants (8 and 12 year olds) by Shahid et al. (2007).

Various lines for future research suggest themselves. We are currently conducting a more detailed analysis of the audiovisual cues in both cultures, to see in which sense the acted emotions are different from the real ones. In addition, we would like to perform cross-cultural perception study looking at the auditory (and the audiovisual) modality. To avoid participants from relying on lexical cues (e.g., "I feel great!") we would have to present the stimuli to judges of a third (non-Dutch, non-Pakistani) culture (cf Barkhuysen et al. 2007).

5. References

- [1] Barkhuysen, P., Kraemer, E., Swerts, M., 2007. Cross-modal perception of emotional speech, *submitted*.
- [2] Darwin, C., 1872. *The expression of the emotions in man and animals*. 3rd edition. Oxford University Press [1998].
- [3] Ekman, P., 1972. *Emotion in the human face*. Pergamon Press.
- [4] Elfenbein, H. and Ambady, N., 2002. On the universality and cultural specificity of emotion recognition. *Psychological Bulletin* 128, 203-235.
- [5] Elfenbein, H., Ambady, N., Mandal, M., Harizuka, S., 2002. Cross-cultural Patterns in Emotion Research, *Emotion*, 2, 75-84.
- [6] Izard C., 1971. *The face of emotion*. Appleton Century Crofts, New York.
- [7] Mackie, D. and Worth L., 1989. Processing deficits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology*, 57, 27-40.
- [8] Mandal, M. K., 1987. Decoding of facial emotions, in terms of expressiveness, by schizophrenics and depressives. *Psychiatry*, 50, 371-376.
- [9] Scherer, K., 2003. Vocal communication of emotion: A review of research paradigms, *Speech Communication*, 40, 227-256.
- [10] Shahid, S., Kraemer, E., Swerts, M., 2007. Audiovisual emotional speech of game playing children: Effects of age and culture, *Proceedings of Interspeech 2007*, Antwerp, Belgium.
- [11] Velten, E., 1968. A laboratory task for induction of mood states. *Behavior Research & Therapy*, 6, 473-482.
- [12] Vogt, T. and André, E., 2005. Comparing Feature Sets for Acted and Spontaneous Speech in View of Automatic Emotion Recognition. In: *IEEE International Conference on Multimedia & Expo (ICME 2005)*.
- [13] Wilting, J., Kraemer, E. and Swerts, M., 2006. Real vs. acted emotional speech, *Proceedings of Interspeech 2006*, Pittsburgh, USA.