On the Role of Acting Skills for the Collection of Simulated Emotional Speech

Emiel Krahmer, Marc Swerts

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

{E.J.Krahmer, M.G.J.Swerts}@uvt.nl

Abstract

We experimentally compared non-simulated with simulated expressions of emotion produced both by inexperienced and by experienced actors. Contrary to our expectations, in a perception experiment participants rated the expressions of experienced actors as more extreme and less like non-simulated ("real") expressions than those produced by non-professional actors.

Index Terms: emotional speech, simulated expressions, acting skills, audiovisual speech

1. Introduction

In research on vocal expressions of emotion, using simulated (acted) expressions has been the preferred way for collecting emotional voice data (Scherer, 2003, p. 232). One important advantage of simulated emotional expressions is that it is generally easier to instruct an actor to display a particular emotion than it is to induce the emotion directly in participants, and, in addition, ethical issues are no stumbling block (which is especially relevant for negative emotions). And as long as the acted, simulated expressions of emotions are representative of non-acted, non-simulated ones there is really no problem. However, so far there is very little concrete evidence that simulated and spontaneous expressions are indeed similar. If anything, there are some suggestions that they actually differ: Wilting et al. (2006), for instance, found that posed expressions are more stereo-typical, and are perceived as stronger than their non-posed counterparts. In a somewhat similar vein, Vogt & André (2005) showed that an automatic emotion recognizer trained on simulated expressions of emotions performs less well when tested on non-simulated emotions and vice versa. In general, it is fair to say that little is known about how simulated and non-simulated expressions relate to each other, even though it is acknowledged that a better understanding of this relation is needed. Scherer (2003, p. 247), for example, states that "obviously, one has to carefully investigate to what extent such acted material corresponds to naturally occurring emotional speech. Unfortunately, so far there has been no study in which a systematic attempt has been made to compare portrayed and naturally occurring vocal emotions."

In addition to this, it seems reasonable to assume that differences in *acting skills* may also have an impact on the quality of simulated emotions. One would hypothesize that more experienced actors would be better in simulating emotions, and thus would produce emotional expressions that are more like natural expressions of emotion. However, researchers using simulated emotional expressions in their studies are mostly not explicit about the skills of their actors, nor about the exact procedure used to elicit simulated expressions.

To find out to what extend acting skills influence the quality of simulated audiovisual expressions of emotion we conducted two experiments. For the first experiment we used a specific adaptation of the Velten (1968) technique, described by Wilting et al. (2006). The basic idea of the Velten technique is that emotions can be induced in participants by letting them read a series of self-referential sentences, that have a progressively stronger emotional content. Since it is a language-based induction method, the Velten technique is particularly relevant for studying the way emotions are expressed through (audiovisual) speech. The original Velten method was used to induce two specific emotions, to wit "elated" (joy) and "depressed" in Velten's terminology. In terms of the dimensional approach to emotions, these two differ primarily along the valence dimension (positive and negative). To find out to what extent simulated expressions differ from non-simulated ones, we added two incongruent conditions to the standard Velten conditions in which participants are explicitly instructed to utter the sentences in a way that is incongruent with their content. They are thus asked to simulate joy while producing the negative sentences, and to simulate somberness while producing the positive sentences. In addition, to find out what the effect is of acting skill, we asked both a group of non-actors and a group of experienced actors to simulate the respective emotions. In a second experiment, we offered spontaneous (no acting) and simulated expressions (of both inexperienced and experienced actors) to a group of judges, who were asked to rate the valence of the expressed emotion. The hypothesis is that simulated expressions from professional actors will be more like non-posed expressions than the posed expressions from non-professional actors are.

2. Experiment 1: Data collection

2.1. Method

Participants 70 speakers participated. Of these, 50 were students and colleagues from Tilburg University (31 females), with a mean age of 27 years. None of these 50 participants was a professional or amateur actor, and none was involved with research on audiovisual speech or emotions. In addition, twenty actors participated, either experienced actors from various theater companies in Tilburg or students in the final year of the Tilburg drama academy. All had between 3 and 25 years of professional experience (*M* = 11.2 years, *SD* = 6.5 years). Ten actors were female, ten male. All participants gave written consent to use their data for research purposes, and none objected to being recorded. Acting participants were randomly assigned to one of the two incongruent conditions.

Materials The sentences used in the various conditions were derived from the original set of sentences used by Velten, consisting of 180 sentences evenly distributed over three conditions (positive, negative and neutral). For this experiment, positive and negative sentences were first literally translated in Dutch, after which they were revised to make sure they were easy to pronounce (see Wilting et al. 2006 for further details). Sentences that referred to ‘specifics’ (e.g., college, parents, religion) were omitted. The neutral sentences (“There is a large rose-growing center near Tyler, Texas”) were replaced with comparable sentences tailored towards the Dutch situation. In the end we selected 40 sentences for each condition. We made sure that the 40 sentences in the positive and negative condition showed the same progression as the original sets of 60 sentences, from neutral (“Today is neither better nor worse than any other day”) to increasingly more emotional sentences (“God I feel great!” and “I want to go to sleep and never wake up.” for the positive and negative sets, respectively), to allow for a gradual build up of the intended emotional state.

Procedure Participants took part one at a time. They were invited to a quiet room, where they were asked to take a seat in front of a desk on which a laptop computer was placed. The laptop was lifted 13cm from the surface so that the screen was more or less at eye level. Right above the screen a digital camera was positioned that recorded the face and upper body of the participants. Participants were told that the camera was only there to check afterwards whether the experimental procedure was properly followed.

Besides the three conditions described by Velten for the induction of real emotions (POSITIVE CONGRUENT, NEUTRAL, NEGATIVE CONGRUENT), two incongruent conditions were added. In one of these, participants were shown the negative sentences and were asked to utter these as if they were in a positive state (POSITIVE INCONGRUENT), in the other, positive sentences were shown and participants were instructed to utter these in a negative way (NEGATIVE INCONGRUENT).

The instructions for the congruent and neutral conditions were a slightly abridged version of the original instructions from Velten. In the instruction phase of the congruent conditions, participants were told that the sentences would represent a “particular emotion” which was not further specified. They were asked to try and “experience” the contents of the sentences, as they were instructed to do in the original Velten method. In the incongruent conditions, participant were told that they would see sentences with a specific emotional content —“sentences radiating positivity and joy” or “sentences radiating somberness and depression”, depending on the condition. They were then instructed to ignore this emotional content, and express the sentences as if they were in respectively a depressed or a joyful state. Other than that, the instructions were exactly the same as those for the congruent conditions. Both the professional actors and the other participants received the same instructions; crucially, the actors were not told that their acting skills were part of the experimental question. In the instructions for the neutral condition participants were merely asked to read each sentence twice, once silently and once out loud. It is important to stress that in none of the instructions for the individual conditions any reference was made to facial or vocal expressions of emotion. Participants were told that the goal of the experiment was to study the effect of emotion on memory recall. The instructions were displayed on the computer screen, and participants were instructed to first silently read the texts, after which they had to read them aloud. This enabled them to practice the experimental procedure. The introduction phase was self-paced.

If the instructions were clear, the experimenter left the room and the actual experiment started. During this phase, the sentences were displayed on a computer screen for 20 seconds, and participants were instructed to read each sentence twice (once silently, then out loud). This phase lasted exactly 800 seconds (40 sentences × 20 seconds), i.e., a little over 13 minutes. During the induction, participants were alone in the room, to avoid presence effects.

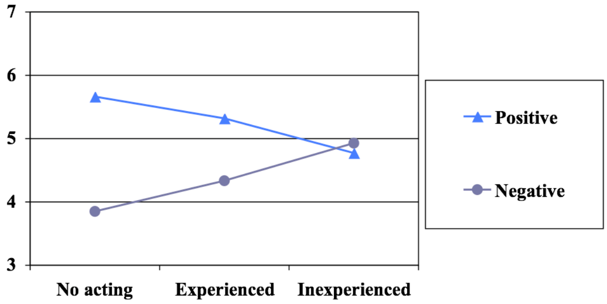

Figure 1: *Self reported emotional state scores for No-acting (non-simulated expressions), Experienced actors (simulated expressions) and Inexperienced actors (producing simulated expressions as well), as a function of valence (Positive, Negative).*

Immediately following this phase, participants had to fill in a short self-report emotion questionnaire (“At this moment, I feel . . . ”) derived from Mackie and Worth (1989) and adapted to Dutch in Krahmer et al. (2004), consisting of six 7-point bipolar semantic differential scales, using the following adjective pairs (English translations of Dutch originals: happy/sad, pleasant/unpleasant, satisfied/unsatisfied, content/discontent, cheerful/sullen and in high spirits/low-spirited). The order of the adjectives was randomized; for processing negative adjectives were mapped to 1 and positive ones to 7. After filling in the questionnaire participants performed a dummy recall test, as this was supposed to be the purpose of the emotion induction. The results of the recall test were not analysed. Finally, participants were debriefed and told about the real purpose of the experiment. They were given a small gift as a token of appreciation.

2.2. Results

Figure 2 shows a number of representative stills of simulated expressions (both from Experienced and Inexperienced actors) and non-simulated expressions in the Positive and Negative conditions. Figure 1 depicts the self-reported emotional state scores from the experienced actors in the positive and negative conditions, and compares them to the scores from participants who did not act and from inexperienced acting participants. The internal consistency of the emotion questionnaire was measured using Cronbach’s α and was very good ($\alpha = .93$).

The self-reported emotion scores were submitted to a 2 (Valence: positive, negative) × 3 (Acting: no-acting, inexperienced-acting, experienced-acting) Analysis of Variance (ANOVA).[1] Overall, participants in the Positive conditions feel more positive afterwards than participants in the Negative conditions, $F(1, 54) = 12.543, p < .001, \eta_p^2 = .188$, while Acting does not reveal a main effect ($F < 1$). However, a significant interaction between Valence and Acting was found, $F(2, 54) = 5.201, p < .01, \eta_p^2 = .162$. This interaction is readily explained by inspection of Figure 1: in the No-acting (Congruent) condition the self-reported emotion scores between

[1] For this analysis, data from speakers in the neutral condition are ignored (they are included in Experiment 2 though).

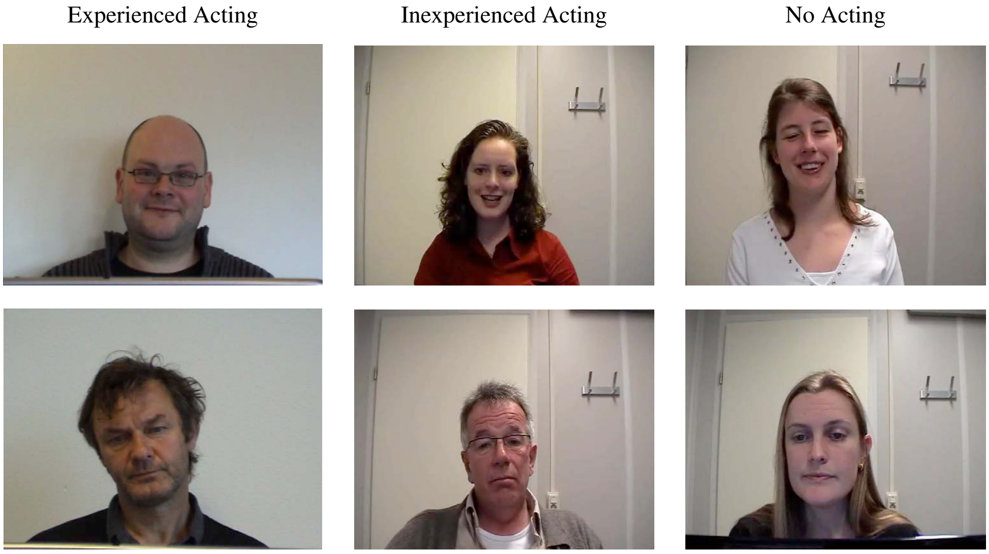

Figure 2: *Representative stills of positive (top) and negative (bottom) expressions of (from left to right) experienced actors, inexperienced actors and non-acting speakers respectively.*

Table 1: *Perceived emotional state scores from adults on a 7-point scale (1 = very negative, 7 = very positive) as a function of condition (standard errors between brackets), with 95% confidence intervals. Two variants of the Incongruent conditions are included, one with Inexperienced Actors and one with Experienced Actors.*

Condition	Perceived emotion (s.e.)	95% CI
Positive Congruent	4.69 (.07)	(4.56, 4.82)
Positive Incongruent Inexperienced Actors	4.70 (.07)	(4.56, 4.84)
Positive Incongruent Experienced Actors	5.71 (.09)	(5.53, 5.89)
Neutral	3.56 (.07)	(3,41, 3,70)
Negative Incongruent Inexperienced Actors	2.89 (.10)	(2.69, 3.09)
Negative Incongruent Experienced Actors	2.40 (.09)	(2.21, 2.59)
Negative Congruent	3.29 (.07)	(3,15, 3.43)

participants in the Positive and Negative Condition are most different, while the differences between Positive and Negative Inexperienced participants are negligible. Interestingly, the scores for the Experienced actors in the Positive and Negative (Incongruent) conditions are almost exactly in between these two extremes: actors in the Positive Incongruent condition, indicate that they feel somewhat more positive at the end of the experiment ($\mathrm{M} = 5.32$, 95% CI $= (4.71, 5.93)$) than actors in the Negative Incongruent condition ($\mathrm{M} = 4.35$, 95% CI $= (3.72, 4.95)$). The crucial question is of course how the simulated expressions from the experienced actors are perceived, especially in comparison with the simulated expression from the inexperienced actors and the non-simulated expressions. This is addressed in Experiment 2.

3. Experiment 2: Perception test

3.1. Method

Participants Forty people participated, of which 20 were female, with an average age of 36.2. None had participated as a speaker in Experiment 1, and none has a background in speech or emotion research.

Materials For each of the speakers in Experiment 1 the final utterance was selected, which arguably catches the speaker at the height of the induced emotion. The resulting 70 stimuli (10 per condition) were cut from just before the participant starts speaking, to just after the sentence was finished, and presented to participants in a vision-only experiment to prevent participants from using lexical cues in their judgment of the stimuli.

Procedure Participants took part one at a time. They were invited into a quiet room, and asked to take place in front of a computer. Participants were told that they would see 70 speakers in different emotional states, and that their task was to rate the perceived emotional state on a 7 point valence scale ranging from 1 (= very negative) to 7 = (very positive). Participants were not informed about the fact that some of the speakers were expressing simulated emotions. The stimuli were offered in one of two random orders, to compensate for potential learning effect. They were preceded by a number displayed on the screen indicating which stimulus would come up next, and followed by a 3 second interval during which participants could fill in their score on an answer form. Stimuli were shown only once. The experiment was preceded by a short training session consisting of three speakers (for which a different sentence was used) to make participants acquainted with the stimuli and task. If all was clear, the actual experiment started, after which there was no further interaction between participant and experimenter. The entire experiment lasted approximately 15 minutes.

3.2. Results

Table 1 summarizes the results. A repeated measures analysis of variance (ANOVA) revealed a significant effect of Condition on perceived emotional state ($F(6, 234) = 360.465, p < .001, \eta_p^2 = .902$). Pairwise comparisons (after a Bonferroni correction) revealed that all conditions were significantly different perceived ($p < .001$), except the comparison between Positive Congruent and Positive Incongruent (by Inexperienced Actors). The emerging picture is surprisingly consistent. Speakers in the Positive conditions are perceived as more positive and those in the Negative conditions as more negative, with neutral precisely in between. Interestingly, in all cases the Incongruent conditions are perceived more strongly than the Congruent ones, albeit that the difference between Positive Congruent and Positive Incongruent (inexperienced actors) is insignificant. And, most interestingly, the stimuli of the Experienced Actors receive the most extreme scores, where the difference with the scores for the Inexperienced Actors is quite substantial, especially for the Positive conditions.

4. Concluding remarks

We have described two experiments, comparing congruent, non-simulated expressions of emotions with incongruent, simulated ones. It was found that non-simulated expressions have a stronger impact on the self-reported emotion scores than simulated expressions; participants that produce simulated sentences feel (close to) neutral afterwards, while participants that produce positive or negative congruent sentences indeed feel more positive or negative. It was interesting to see that the self-reported scores from the professional actors were almost exactly halfway between the scores of the non-acting participants and the non-professional actors.

We hypothesized that simulated expressions of professional actors would be more realistic (i.e., more like non-simulated congruent ones) than those of non-professional actors, and this was tested in Experiment 2. However, it turned out that, contrary to our expectations, the expressions of the experienced actors were perceived as even more extreme than those of the participants without an education in and professional experience with acting. Naturally, it can be claimed that if the actors would be trained using, say, the Stanislavski method or if they had a background in method acting they might display more subtle expressions (e.g., Scherer, 2003; Marsella et al., 2006), or alternatively that expressions that are elicited using extensive scenarios (Enos & Hirschberg, 2006) would be more realistic. But surely one would expect that expressions from experienced actors would at least go some way in the more realistic direction, which is clearly not what we found. In general, the findings suggest that the simulated expressions of both experienced and inexperienced actors in our experiments are more intense than the non-posed ones (hence the more extreme scores in Experiment 2). Inspection of the data suggests that non-posed expressions often do not consist of the "complete" stereotypical expression associated with joy (pronounced smile, raised brows) or somberness (frown, mouth corner pulling), which is in line with the claims from, for instance, Horstman (2002) that people not frequently display entire stereotypical expressions spontaneously. Finally, it is worth pointing out that our findings are consistent with the few studies which have directly addressed emotions in acting. In particular, Konijn (2000) found that actors typically indicate that they do not feel the emotions they are instructed to display (although they may experience task emotions related to the acting itself).

Arguably, one limitation of Experiment 2 is that it is based on facial expressions only. Presenting the recordings to participants (judges) in an audiovisual format is complicated since the lexical material of the sentences is a give-away clue for the emotional state of the speaker. However, Barkhuysen et al. (2008) presented a selection of the stimuli collected in Experiment 1 in three conditions (audio-only, vision-only and audio-visual) to Czech participants (not speaking Dutch). This revealed similar results as described in Experiment 2 here (simulated expressions perceived stronger than non-simulated ones) for all three modalities, although the differences in the visual condition were more pronounced than those in the auditory one.

Clearly, showing conclusively how simulated and non-simulated emotional expressions relate to each other, cannot be done with a single series of experiments. It would be very interesting, for instance, to find out what the results would be for different acting methods, for non-read speech, and for other emotional states than those under study here.

5. Acknowledgements

We would like to thank Lennard van de Laar, Marieke Rikken, and Janneke Wilting for their help with these experiments, Harold Miesen and Carel van Wijk for statistical advice, and to Diederik Stapel, Kirsten Ruys and Emmett Velten for useful discussions.

6. References

[1] Barkhuysen, P., E. Krahmer & M. Swerts (2008), Cross-modal and incremental perception of emotional audiovisual speech, *Language and Speech*, accepted pending minor revisions.

[2] Enos, F., & Hirschberg, J. (2006). A framework for eliciting emotional speech: Capitalizing on the actor's process. In *LREC 2006 Workshop on Corpora for Research on Emotion and Affect*, Genova, Italy.

[3] Horstmann, G. (2002). Facial expressions of emotion: Does the prototype represent central tendency, frequency of instantiation, or an ideal? *Emotion, 2*, 297-305.

[4] Konijn, E. (2000). *Acting emotions: Shaping emotions on stage*. Amsterdam University Press.

[5] Krahmer, E., J. van Dorst and N. Ummelen (2004). Mood, persuasion and information presentation, *Information Design Journal, 12*, 40–52.

[6] Mackie, D. & L. Worth (1989). Processing decits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology, 57,* 2740.

[7] Marsella, S., Carnicke, S., Gratch, J., Okhmatovskaia, A., & Rizzo, A. (2006), An exploration of Delsarte's structural acting system. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA)*. Marina del Rey, CA.

[8] Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227256.

[9] Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research & Therapy, 6*, 473-482.

[10] Vogt, T., & Andr, E. (2005). Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In *Proceedings of IEEE International Conference on Multimedia & Expo (ICME 2005)*, Trento, Italy.

[11] Wilting, J., E. Krahmer & M. Swerts (2006), Real vs. acted emotional speech, *Proceedings of Interspeech 2006*, Pittsburgh, PA, USA.