

A Framework for Eliciting Emotional Speech: Capitalizing on the Actor’s Process

Frank Enos

Department of Computer Science
Columbia University, New York City
frank@cs.columbia.edu

Julia Hirschberg

Department of Computer Science
Columbia University, New York City
julia@cs.columbia.edu

ABSTRACT

This paper offers an approach and a theoretical framework for eliciting emotional speech using actors. The framework is developed by connecting the goal-based model of emotion proposed by Abelson [1], the work of appraisal theorists, and an approach to the actor’s technical process widely used in the professional theater and taught in modern conservatories. In doing so, we hope to address some of the difficulties currently encountered in the use of acted speech in emotion research.

Author Keywords

Emotion, elicitation, acting, appraisal theory.

ACM Classification Keywords

I.2.7 [Artificial Intelligence]: Natural Language – *Speech Recognition and Synthesis*

INTRODUCTION

Numerous studies of emotional speech have employed acted speech with varying degrees of success. We believe that the utility of this approach to elicitation can be augmented by a more thorough understanding of the actor’s process and how it relates to emotion. Campbell [3] gives a thorough accounting of the problems presented by the use of acted speech. He details the concern that such speech may be limited in dimension, and, more subtly, that in social interaction, expression of emotion may be unintentional. In contrast, in most studies that use acted speech, the expression of a particular emotion is both intentional and overt. Indeed, many such studies simply instruct the actor to utter a given text “with emotion X” (for a listing of such studies, see [11]; a list of databases can be found in [6]). Notable exceptions include those studies that employ the ‘scenario approach’ such as [2] and a study currently underway by the Geneva Emotion Research Group [Bänziger, personal communication]; Scherer et al. [10] provide a list.

Even in studies that employ scenarios, however, there is usually overt instruction to the actor with respect to specific emotions. We argue that directing the actor to focus on the goal of encoding a particular emotion works at cross-purposes with the advantages conferred by the trained actor’s process. Consider a play in which an actor attempts to convey sorrow by forcing himself to weep. Such a ploy is often unconvincing, and transparent even to a naïve observer. Most importantly, it is generally inconsistent with human behavior. This is an extreme form of the sort of arti-

ficial manipulation entailed in many emotional speech elicitation paradigms. It assumes that the actor has been trained – and has achieved the capacity – to manipulate directly the physiological mechanisms involved in the expression of natural emotion. In most cases, this is simply not true.

Preliminary assumptions

For the purposes of this paper, emotion is understood to be defined in the following complementary ways:

1. ‘Reaction to significant stimulus events that impinge on organismic equilibrium or that change the organism-environment relationship.’ [9]
2. ‘Valenced reactions to events, agents or objects, with their particular nature being determined by the way in which the eliciting situation is construed.’ [8]

Further, the consideration of acting given here is primarily concerned with approaches to stage acting, in particular, acting in the tradition of the American and British stage. This is because in theater, as in daily life, the perception of emotion on the part of an observer or interlocutor is to a degree unmediated by other forces. In film or television, the director, via the camera and editing, makes crucial decisions with respect to what is observed [McEleney, personal communication], and thus the conveyance of emotion in film depends heavily on a number of factors external to the actor and the observer. Although the stage director controls many elements of the interaction, the contact is ultimately live and direct. Finally, it is not suggested that all actors work in the way described here. The approach described is, however, founded on six years of conservatory training and fifteen years of professional experience on the part of one of the authors, and we have found it to be one effective means of accomplishing the actor’s task.

TECHNICAL ASPECTS OF ACTING

We begin with a personal observation: an inexperienced or poorly-trained actor who is having trouble conveying a scene convincingly can often readily report what he or she is feeling. In contrast, a convincing actor will often report what he or she is doing. The latter is consistent with the approaches advocated by many respected teachers of acting, for example Michael Chekhov [4], the faculty of the Brown/Trinity Rep Consortium, and at its’ root, Konstantin Stanislavski [14]. In these approaches, (in Chekhov’s [4] terminology) the actor develops an ‘action’ (possibly an

intention; sometimes literally an action) for each segment of the script, and, working with the director, develops elements of character (personality) and atmosphere (context) in which to pursue the action. We argue below that this approach has the potential to simulate the appraisal process. This happens in terms of the personality and background of the character, and in terms of the events, agents, and context of the given scene. In turn, this simulation allows the mechanisms of emotional expression — facial, vocal, physical — to engage naturally. This process is complemented by aspects of the actor's training that develop physical and mental responsiveness to emotion inducing-stimuli.

We do not suggest that all approaches to acting are reducible to this approach. We do argue that this approach may lead to more viable and natural elicited emotional speech than has been previously obtained from trained actors.

What (some) actors do

A widely accepted claim in acting is that “Good actors try to do what people do” [McEleney]. What then distinguishes theater from the events of our daily lives? We contend that the appeal of theater is found in the opportunity to observe ‘what people do’ in extraordinary situations. These situations may be extraordinary because of momentous events such as the shipwreck that precipitates the events of Shakespeare’s ‘Twelfth Night’; because of an unusual context, such as the enchanted island of ‘The Tempest’; or because of the emotionally charged relationships of the characters. A good play often combines these characteristics. This claim can thus be reframed: Good actors, on stage, try to do what people do under extraordinary circumstances. In the best case, such portrayals necessarily generate emotion, since ‘human behavior in a high stakes situation’ is virtually synonymous with the eliciting circumstance described in Definition (1) of emotion above. Crucially, the intensity of emotion is modulated by what is at stake.

Acting and goals

What exactly is meant here by human behavior? Certainly not ‘to have emotion X’. We claim, for example, that it is not a human activity, in the sense of a primary, intentional activity, to be sad. Display rules [7] may dictate that one appear sad when one is actually happy (e.g. a rival fails spectacularly). The operative word, though, is ‘appear’, since the true emotion experienced is that of happiness; the intentional activity in this case is ‘to appear X’.¹

Novice actors often confuse behavior with ‘pretend’ behavior [Berenson]. The skilled actor does not pretend to pick up a pistol from the stage floor, he simply picks it up. The manner in which he picks it up is dictated by his understanding of what he, the actor, would have to do in order to do what the character does. In this version of what

Stanislavski called the ‘Magic If’ [13], the ‘pretend’ takes place in the actor’s preparation. As he clarifies the imagined circumstances and goals, it becomes increasingly easier to commit to the action. In turn, emotion arises from the interaction of goals, context, and obstacles, which become increasingly believable to the actor via rehearsal. This obviates the need to manipulate the display of emotion.

The trap of focusing on the emotion is often described by actors as ‘playing the result’ rather than ‘playing the action’ [Berenson, personal communication]. In focusing on the result, the actor is asking the question ‘what would I sound like?’ instead of ‘what would I *do*?’ The actor tries to manipulate the body, face, and voice in order to indicate based on some preconception. Again, this is not human activity: the actor engages in behavior only an actor would perform.

EMOTION, GOALS, AND APPRAISAL

Abelson [1], writing on cognitive consistency theory, proposes what he calls a functionalist model of emotion. His model contains the following components: Goals, Actions (of the subject), Causal Instrumentality (other agents or events), and Outcomes. Emotion arises as a consequence of the goal relationships in a situation, possibly moderated by expectation. For example, disappointment arises when, given a goal, action, and the influence of some agent, a positive outcome is expected, but the result is instead negative. Likewise, frustration arises when, given the individual’s goals and actions, a positive outcome might be expected, but is thwarted by the interference of some agent.

It further seems straightforward to relate Scherer’s [9] four major classes of appraisal criteria to Abelson’s goal-based model, and consequently to the acting approach we offer. These criteria are (briefly): (1) intrinsic characteristics of objects or events; (2) the significance of the event for the individual’s needs or goals; (3) the individual’s ability to influence or cope with the consequences of the event; (4) the compatibility of the event with social or personal standards. Abelson’s model is equally compatible with other extant work in appraisal theory, for example, Clore’s [5] Immediacy Principle: ‘Affective feelings tend to be experienced as reactions to current mental content.’ And one reading of this principle suggests that the more thoroughly the actor commits to her beliefs with respect to the character, the closer she comes to producing actual emotion.

Although not fully developed here, this connection of appraisals with goals may help to clarify why a goal-based approach could be effective in producing more realistic emotion than is usually achieved by actors in the laboratory. Since some appraisal criteria — such as beliefs regarding the characteristics of objects or the significance of events with respect to goals — may be influenced by imagination, the actor’s work may simulate appraisal criteria in a manner that allows emotion to engage and arise naturally.

TWO SUGGESTED APPROACHES

The task of designing an emotion elicitation experiment, is, in a sense, to work backwards. One can deconstruct the

¹ Granted, it is not an uncommon goal ‘to be happy’, but this entails a more existential scope than does the corresponding emotional state.

desired emotion into component parts – goal or intention, context, and text – which the actor then ‘reassembles’ into emotion, spoken and otherwise. We offer two approaches to eliciting emotion in this fashion. The first revisits methods used by other researchers, but with a focus on the component parts described here. The second makes direct use of the process and materials most familiar to actors.

The Scenario Approach

A number of studies take this approach (e.g. [2,10]), and it is highly compatible with the methods proposed here. This is even more true for those paradigms that develop scenarios in concert with the actors, possibly via improvisation, as in a recent data collection effort by the Geneva Emotion Research Group [Bänziger, personal communication]. We add to approaches already in use the idea of reverse-engineering the desired emotion into the components most useful to actors. Such an eliciting scenario is composed of:

1. A description of the character.
2. A clear description of the situation, including obstacles to the goal, and what is at stake.
3. A (single) goal for the actor.
4. A text that is meant to accomplish the goal, regardless of whether the specific lexical content embodies that goal.
5. Description(s) of additional character(s) involved, if any.

Below is an example scenario for the emotion anger:

1. The speaker’s name is June. She is the mother of three children living in a small American town in the state of Pennsylvania. She works during the day as a waitress in an inexpensive restaurant. Most nights, she cleans the offices of a man who is a customer of the restaurant. Most nights, the offices do not need to be cleaned. The owner has given her the job because he feels sorry for her. She knows this. One year prior to the scenario, her husband Bill went to the corner store to buy a lottery ticket and never returned. Her family lives in Arkansas; she left Arkansas with Bill seven years earlier to move to New York, where he planned to work as a professional poker player. Pennsylvania was the closest they came to New York. Bill worked for a time in his second cousin’s an auto repair shop.
2. June has just returned home at midnight on a week night from cleaning offices to find Bill sitting in the living room playing with the children, cigarette in his mouth and can of beer in his hand. The children worship him like a conquering hero returning from battle.
3. To make her husband leave her home.
4. ‘Don’t even speak to me!’
5. Bill, as described above, smug and unapologetic.

The level of detail in this example may be *less* than that desired in an actual elicitation scenario. It is important to understand that the average professional stage production rehearses for six weeks or more. In such a production, good actors, often assisted by the playwright, director, and other actors, develop a very specific picture of the life and world of their characters. This may include details not superficially relevant to the action of the play: what foods the character dislikes; what magazines she reads; what she

wears to sleep. The more the actor ‘knows’ about the character, the more readily she can commit to her behavior.

The five elements can be viewed as an equation, the output of which is human behavior, including emotion [Berenson]. The work of the researcher, assisted by a skilled director, is to tune the parameters of this equation in order to elicit the desired result. This should happen *without mention to the actor of the desired emotion*. The charge to the actor is to accomplish some action by uttering the text in such a way that it effects the goal; the actor’s focus must be on action.

If, for example, the result is not sufficiently angry for the purposes of the elicitation, rather than focusing the actor on this result, the director might raise the stakes: ‘Imagine now that Bill has been gone for three years.’ Or that Bill laughs and begins to speak. Perhaps the goal is cold anger: ‘Imagine that Bill has been gone for 10 years.’ Campbell [3] raises the valid concern that acted emotion is too overt. Perhaps June wishes to make Bill leave without upsetting the children – how does she effect this? The subtext of anger will remain, but the focus is on what June is *doing*; the actor allows the emotion to take care of itself. A primary tool for doing so is specificity [McEleney]. The rehearsal process that fills out the life, the world, and the personality of the character is a process of making clear, specific choices. A scenario can be filled out by embedding the target utterance in a scene, or actors can improvise so that they can more thoroughly commit to the events.

We suggest two possible methods for a priori validation. First, scenarios can be rated for emotional content by a group of judges who can be asked to determine the emotion they would likely experience as the character(s). In this way, one can have some assurance beyond intuition that the given scenario is likely to elicit the desired emotion. Second, a more data-driven approach to scenario development might be taken (as in e.g. [10]), in which scenarios are constructed as above but based on an existing database of actual subject experience. One such a database is the International Survey On Emotion Antecedents And Reactions (ISEAR) [12], which details the emotional experiences of thousands of cross-cultural subjects.

The script approach

A second proposed method of elicitation seeks to make direct use of the actor’s training and skills. In this approach, existing, well-crafted scripts are used, in their entirety or in the form of individual scenes, as the scenarios of elicitation (scenes may be validated a priori as above). If actual language is to be used, it is preferable to select a script that uses contemporary language; certainly language that is likely to be familiar to any subjects who may label the speech. If pseudo-linguistic content (see below) is to be used, the script may be selected from any that are approachable by the actors employed, since it will serve only to develop the actions and events that elicit the emotion. Development of scenarios proceeds as previously described: the five points mentioned above are developed with respect to each scene of interest, and the text is chosen from

the play itself. This approach has the advantage that a well-crafted play is literally designed for the actor's methods. The characters, events, and goals will present themselves, with the help of a director, through the process of rehearsal. Take for example a scene from *Romeo and Juliet* of Shakespeare (chosen for its familiarity; see note below on language). The goal here is to elicit from Romeo the emotion 'despair' as he conveys to Friar Lawrence his dismay at having been banished for killing the character Tybalt.

*Hadst thou no poison mixed, no sharp-ground knife,
No sudden mean of death, though never so mean,
But 'banished' to kill me? – 'banished'?*
O friar, the damned use that word in Hell; [13]

Here the script provides a goal (to be with Juliet), a context (banishment, never to see her again), and a well-developed atmosphere and set of characters. Tokens can be selected from the text, and the actors coached as desired.

Both these approaches are compatible with elicitation scenarios that employ pseudo-linguistic content, but it is crucial to avoid the actor's imposition of preconceptions with respect to the 'sounds' of emotions. Again, the solution is to approach the speech from the standpoint of intention or goals. With the archaic language of Shakespeare, the actor develops an intention and a set of internal and external circumstances so specific that she could only express them in the language of the script [McEleney]. Analogously, the actor must imbue pseudo-text with an intention so specific that the lack of semantic meaning is inconsequential.

CONCLUSIONS

This paper has described an approach to eliciting acted emotional speech that makes use of the trained actor's process. This approach avoids conveying specific emotional labels to the actor in favor of providing goals and contexts in which to play out certain actions. It is suggested that in this manner the actor can engage the mechanisms of appraisal, and thereby produce emotion in a more natural way. The authors have not addressed the process of post-hoc validation of the elicited emotion. It should also be mentioned that this approach assumes the involvement of skilled professional actors; the outcome of using this approach with college or amateur actors is likely to be disappointing. Finally, we think two areas merit further consideration. First, this approach may also be useful in studies of facial expression or body posture (e.g. [15]). Finally, we find the interaction between goals and appraisal interesting, in the context of acting and otherwise.

ACKNOWLEDGMENTS

We thank Tanja Bänziger for information about her current work; and Marc Schröder for comments on some of the ideas presented here. We are grateful to Stephen Berenson and Brian McEleney of the Brown University / Trinity Repertory Theater Consortium for helpful conversations

about this work. This research was funded in part by NSF IIS-0325399 and the Department of Homeland Security.

REFERENCES

1. Abelson, R.P. What Ever Became of Consistency Theory? *Pers. and Social Psych. Bull.* 9(1), 1983, 37-54
2. Banse, R. & Scherer, K.R. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), (1996), 614-636.
3. Campbell, N. Databases of emotional speech. *Proc. ISCA Workshop on Speech and Emotion*, (2000) 34-38.
4. Chekhov, M. *To the Actor*, Harper & Row, NYC (1953).
5. Clore, G. L., Gasper, K., and Garvin, E. Affect as Information, in J. P. Forgas (ed.), *Handbook of Affect and Social Cognition*, Lawrence Erlbaum Associates, Inc., Mahwah, NJ (2001).
6. Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. Emotional speech: towards a new generation of databases. *Speech Communication*, 40, (2003) 33-60.
7. Ekman, P. Universals and cultural differences in facial expressions of emotion. In J. Cole (ed.), *Nebraska Symposium on Motivation*, 1971. University of Nebraska Press (1972), 207- 283.
8. Ortony, A., Clore, G. L., and Collins, A. *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, UK, (1988).
9. Scherer, K. R. Appraisal theory. In T. Dalgleish, & M. Power (Eds.). *Handbook of Cognition and Emotion* (pp. 637-663). Chichester, (1999).
10. Scherer, K. R., Banse, R., & Wallbott, H.G. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), (2001), 76-92
11. Scherer, K. R. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, (2003), 227-256.
12. Scherer, K. R., & Wallbott, H.G., Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66, 310-328, (1994).
13. Shakespeare, W. *Romeo and Juliet, The Riverside Shakespeare*, Houghton Mifflin, New York, (1974).
14. Stanislavski, K. *An Actor Prepares*, Hapgood, E.R., trans., Theater Art Books, New York, 1971.
15. Wallbott, H. G. Bodily expression of emotion. *European Journal of Social Psychology*, 28, (1998), 879-896.
16. Bänziger, T. personal communication, March 30, 2006.
17. McEleney, B. personal communication, March 21, 2006.
18. Berenson, S. personal communication, March 26, 2006.