Making space for emotions in speech research: a methodological consideration

Thunberg, G. C.

journal: TMH-QPSR
volume: 37
number: 2
year: 1996
pages: 093-096

http://www.speech.kth.se/qpsr
Making space for emotions in speech research.  
- A methodological consideration

Gunilla C. Thunberg  
Dept of Linguistics, Stockholm University

Abstract
Research on emotional expressions in speech is mostly built on the assumption that human emotions, the way they are experienced as well as the way they are expressed, can be readily sorted into categories like: anger, joy, fear, and sadness. These four categories are often referred to as being our basic or primary emotions.

However, in experimental situations listeners frequently find it difficult to discriminate some of the presented emotional stimuli; e.g. intended joy tends to be confused with anger. This suggests that, although those two expressions represent quite opposite emotions, they probably share at least one common feature.

Furthermore, it suggests that we would need a more complex way to describe emotional speech, in order to make analytic progress. The present paper is a first attempt to outline a multi-dimensional model for the description of emotional speech, thereby allowing features to overlap between the traditionally basic emotional categories.

A complex introduction
Emotions are complex phenomena, the way they are experienced as well as the way they are expressed; let alone the way they are conceived by a communicative partner.

Any kind of situation demanding an emotional response will evoke complex neuro-physiological processes, including hormonal activities, resulting in some kind of complex behavior displaying the nature of the experienced emotion(s). The kind of activity involved will of course vary with the nature of the triggering situation. In spontaneously expressed emotions the means for expression will comprise the entire body, engaging all the signalling systems available, while as in the more deliberately controlled expressions we may choose the means that would best serve our purposes.

Intentional influence on the message
In most cases we have more or less obvious intentions when communicating our emotional messages to other people. Usually there is some kind of desire or intended achievement involved, which will also be reflected in our choice of expressive features.

Contextual influence on the message
There is always a reason behind an emotional experience. A triggering situation might have varying impacts depending on our previous experience, and the resulting reaction will provide some information about that background.

Be careful not to kill the message
Influence from context and intention, altogether, will govern the way we choose to express ourselves, and we may hypothesize that those two conditioning factors are indeed essential for communication as a whole.
From that it would follow that emotional expressions are not messages per se, but merely constitute the tools we use to convey those messages. What happens then, if the tools are dull or if every other saw tooth is missing? An incomplete message will increase the risk for misinterpretation and confusion.

As an example: imagine a smiling face without the friendly sparkle in the eye. The smile will be reduced to just a grinning grimace around the mouth. Such an expression will be interpreted as something completely different from a friendly smile - maybe even as its direct opposite: rejection and hostility. The somewhat drastic heading above was chosen to imply that some misinterpretations may actually prove to be fatal!

Maybe not so, in a normal experimental setting, dealing with research on emotional speech. But still, one cannot help wondering if the above given example might have a bearing on the fact that joy is quite frequently confused with anger under experimental conditions.

On the importance of being earnest

The confusion probably derives from one - or more - commonly exposed features, e.g. excitement. If the appropriate supporting information is lacking; or maybe is not strong enough to provide a clear direction (i.e. positive vs. negative), the interpretation would be based on chance rather than a properly conceived impression. Under such circumstances, listeners will be subjected to heavy contextual impact pertaining to the experimental situation.

As always, stimulus order as well as given instructions will be of great importance for listeners' judgments, but even more so if the test material has got deficiencies. What is crucial, however, is our choice of the appropriate parameters for listeners to judge. If a chosen parameter or the way it is labelled is unclear or ambiguous, this will also yield confusion, no matter how much information can be provided by the stimuli. Picking the appropriate parameters is yet another necessary condition for the assessment of those features that are commonly shared, and those that would be used in distinguishing emotions.

Don't worry, be happy!

Both sadness and joy could make us cry. Just observing the tears would not give you the whole picture. In order to elicit valid emotional expressions to be recorded for experimental purposes, informants must be given sufficiently complex instructions. “Please, cry when you say this sentence!” obviously would not suffice to elicit a genuine expression, with a built in motivational context.

Just like the above phrase: ‘don’t worry, be happy’ is a mere cliche; an empty message, and a poor piece of advice; an instruction given to a speaker may sometimes be overly simplified, based on expectations for certain superficial effects, rather than any well founded ingenious expressions. The resulting material, then, runs a great risk of turning out as overly simplified, too; as if it were synthetic and undersampled (cf. ‘every other saw tooth missing”).

If we are fortunate, we may of course have hit the right key, i.e. the feature that would distinguish an expression from another. But how are we to know, for certain, if we have disclosed the whole truth, without having a profound knowledge about the complexity in emotional expressions?

A first conclusion

To conclude this introductory section, we argue that expressive complexity should be the main objective of research on emotions in speech. Rather than trying to apply the expressed emotions to differently pre-labelled categories, we should be looking more closely into the ways they are related. This calls for a multidimensional model for description, and in the following section we will outline one possible method to really make space for emotions in speech research.

Suggested method

The method reported in this section is being used in an ongoing study, with the aim to develop a reliable, multi-dimensional model for the description of emotional speech. It can be divided into three steps: elicitation of emotional expressions; experimental design; and possible outcome.

Elicitation of emotional expressions

The informant(s) are asked to produce emotionally varying versions of a single sentence, elicited by triggering phrases that will serve to induce diverse contextual conditions for each elicitation. Both the triggering phrase and its consecutive utterance is recorded, using hi-tech equipment (sound proof environment and DAT recorder).
The present study is based on the Swedish sentence: “det har är bara en mening” ('this is just a/one sentence'). The triggering phrases are: e.g. 'I thought this would take all night to correct, but....' (supposedly inducing a positively surprised expression); or 'I told you to write one whole page, but....' (somewhat aggravated); etc. The recorded material thus can be made to display even varying degrees of emotional realizations, ranging from slightly affected through heavily provoked responses.

Experimental design

A randomly ordered test material is put together, containing only the elicited second portion of each utterance, without the triggering phrases. It is presented via earphones to a panel of listeners, who are asked to make four separate judgments for each elicitation. To be able to perform this task, they will be exposed to each stimulus four times in a row, at approximately 3 sec intervals. The first three judgments are of psychometric nature and will be marked along visual analogue scales (VAS). The fourth answer to be given, is a brief characterization of the emotional message, written in the subject's own words.

Visual analogue scaling

The VAS is a method frequently being utilized in the social and behavioral sciences, for measuring all kinds of subjective phenomena. Apart from its gains within medical research, it has also proved beneficial for clinical purposes, such as for the estimation of experienced pain in patients.

VAS scales may differ in fashion. However, a non-graded horizontal line with a length of minimum 10 cm is a commonly used variety, where the scores can be obtained by measuring the distance, usually in millimeters, from one end of the scale to the subject’s mark. The line should have labelled 'anchors' at both ends, defining the extreme points for some determined parameter.

The VAS described above, is exactly the one used in the present study, where one of the chosen parameters is ‘negative vs. positive’, as shown in fig 2. If a second parameter, then, is chosen to be, e.g. ‘minus vs. plus excitement’, we may get similar scores for that one, for expressed joy and anger. However, the first parameter might be judged differently for those two, which could indicate its significance for discrimination in this case.

Figure 2. VAS line with anchors at both ends, defining the parameter to be judged. Note that a real size VAS should be at least 10 cm long.

Choice and definition of parameters

The parameters to be judged are chosen to represent some valid perceptual dimensions. They must be possible to discern in a complex expression, and they should be thought to have a bearing on the interpretation of the message.

The choice of terms to define the parameters is yet another delicate matter. The terms should not be of the kind referring to traditional emotional categories like ‘happy’, ‘sad’, etc., since that would produce answers working counter to our aim at analytic decomposition of the emotional complexity. On the other hand, a term like ‘intensity’ should be avoided, since it would be likely to bias, at least non-naïve listeners, towards the terminology used in acoustic metrics.

Characterization of the message

The written comment, following the psychometric judgments, may serve as a checking point, in more than one way. Above all, it would tell us whether the perceived message actually matches the one intended. Should this not be the case, we might hypothesize, e.g., that not enough contextual information was present in the elicited utterance.

One may also be able to estimate the degree of contextual impact due to stimulus order, by comparing those comments made on the same stimulus, presented a number of times in varying contextual environments.

When given the opportunity to describe the emotional content of the messages in their own words, subjects tend to favor a more complex characterization. This in itself lends support to the notion of complexity. The freely chosen words may also serve to indicate the feature(s) perceived to be predominant in the message.

Possible outcome

Data collected in the way described here should be carefully evaluated before getting down to further analyses. The entire test procedure may have to be repeated, to make sure the chosen parameters and their defining terms are appropriate.

Once we get the whole picture; i.e. when we arrive at a multi-dimensional description,
allowing features to overlap between the traditional categories; we may proceed to analysing the speech material acoustically. If the descriptive model is reliable, we should be able to find acoustical correlates to the perceived dimensional differences.

There may not be any simple match between the concept of a perceptual dimension and an acoustically well defined trait, that can be readily quantified. Furthermore, the aspect of voice quality as an expressive means, must also be taken into consideration.

Concluding discussion

In the present study, we are now in the process of evaluating the test procedure.

Some of the terms defining a chosen parameter have been rejected, since apparently they were not clear enough. Listeners were asked to judge emotional strength along a 'weak vs. strong' VAS line. However, these terms were interpreted differently among subjects, as could be inferred from the very large scoring deviances. Alternative terms are being tried out.

Another issue, to be dealt with, is the question whether the written characterizations might become biased, by the mere fact that subjects are asked to make judgments repeatedly for each stimulus, before making the overall comment. This might induce an analytic approach in the overall characterization, not compatible with a spontaneous interpretation. Or, it may well be the other way around. Some of the subjects, so far, have reported on a tendency to start considering the overall impression, before making their judgments on the defined parameters.

Several factors have to be taken into further consideration, in order to develop a reliable model for the description of emotions in speech. Suggestions in benefit for its construction are invited. For communication, preferably use this address: gunilla.c.thunberg@ling.su.se.

References


