# Detecting Emotional Involvement in Professional News Reporters: An Analysis of Speech and Gestures

**Maria Grazia Busà, Alice Cravotta**
University of Padua
Via B. Pellegrino, 26
35137 Padova
E-mail: mariagrazia.busa@unipd.it, alice.cravotta@unipd.it

**Abstract**

This study is aimed to investigate the extent to which reporters' voice and body behaviour may betray different degrees of emotional involvement when reporting on emergency situations. The hypothesis is that emotional involvement is associated with an increase in body movements and pitch and intensity variation. The object of investigation is a corpus of 21 10-second videos of Italian news reports on flooding taken from Italian nation-wide TV channels. The gestures and body movements of the reporters were first inspected visually. Then, measures of the reporters' pitch and intensity variations were calculated and related with the reporters' gestures. The effects of the variability in the reporters' voice and gestures were tested with an evaluation test. The results show that the reporters vary greatly in the extent to which they move their hands and body in their reportings. Two gestures seem to characterise reporters' communication of emergencies: beats and deictics. The reporters' use of gestures partially parallels the reporters' variations in pitch and intensity. The evaluation study shows that increased gesturing is associated with greater emotional involvement and less professionalism. The data was used to create an ontology of gestures for the communication of emergency.

**Keywords:** Non-verbal behaviour, gestures, speech analysis, emergency communication.

## 1. Introduction[1]

Studies have shown that in the communication of affect (i.e., emotional involvement), words account for under 10% of the meaning exchanged, while around 40% of the meaning is transmitted through paralinguistic features of the speakers' voice (e.g., pitch, voice volume), and about 50% through body language (Mehrabian & Wiener, 1967; Mehrabian & Ferris, 1967; also reported in Mehrabian, 1972). Different modalities lend themselves to representing certain kinds of information better than others. For example, the hands express shapes better than speech; the face expresses attitudes better than words.

Investigating how people communicate emotion through their use of non-verbal language is important on theoretical grounds as well as for extracting data that can be used in information analytics systems. Such systems must take into account that communication involves different modalities (written, spoken, non-verbal). This is of particular importance today since the social media provide an increasing amount of content through audios and videos.

While it seems feasible to extract multimodal data from audios and videos, there is a lack of a solid body of research on the relation between acoustic features of speech and gestures in emotive communication that would help getting information about the emotive state of the speaker.

Investigations of emotive vocal behaviour (Scherer, 1986, 2003; Juslin & Scherer, 2005; Juslin & Laukka, 2003) have associated certain changes in voice acoustic patterns to basic emotive states. As far as body and gestures are concerned, research has focussed more on facial expressions (Ekman & Friesen 1977; Ekman, 1993; Ekman et al., 2013) than on the expression of emotions through body postures and gestures. Facial expressions have been shown to be universally associated to a set of basic emotions, while gestures and body posture were left apart in defining the quality of an affective state (De Gelder, 2009). But there is evidence that emotions are manifested with a synchronized response incorporating physiology, speech, facial expressions, modulations of posture and gestures, affective vocal behaviours and actions.

The present study was carried out within the EU FP7 Security Programme sponsored Slandail project. The aim of the project is to make ethical use of the information available in the social media to enhance the performance of emergency management systems. As part of this project, research is carried out aimed at extracting and integrating text, image, speech and non-verbal data from the social media.

In this framework, the analysis of integrated speech and gesture data can be used (i) directly, as indicators of the speakers' emotional involvement as an effect or a reaction to a disaster and, (ii) indirectly, to obtain information about the gravity of the disaster.

This study aims to examine whether speech and non-verbal language can be used to extract data on speakers' emotional involvement. For this purpose, it investigates a corpus of videos on journalists reporting on natural disasters. The study consists of (1) a qualitative analysis of the gestures performed by the reporters in their reportings; (ii) an acoustic analysis of some speech characteristics (pitch, intensity) of the reporters' voices; (iii) an evaluation of the relation between the reporters' speech characteristics and their levels of body dynamism; (iv) an evaluation of the effects of the reporters' speech

[1] The research leading to these results has received funding from the European community's Seventh Framework Programme under grant agreement No. 607691 (SLANDAIL)

and gestures on the audience public. The study also draws the ontology of the two gestures that appear most commonly in emergency communication.

## 2. Detecting Frequent Gestures in the Reporting of Disasters: A Qualitative Study

Journalists reporting on emergency situations must appear professional and not emotionally involved with the event they are reporting. However, while they may manage to do so with their verbal language (choice of words, discourse structure), they may communicate emotional involvement with their body gestures. It is in fact possible that the more they are involved in the situation they are reporting the more this involvement will leak out from their body movements.

This led to formulate our first set of hypotheses, that is: 1) reporters convey emotional involvement in a disaster situation through their non-verbal language; 2) when they are emotionally involved in the reported event, reporters will tend to use some gestures more frequently than others.

To test these hypotheses, the study reported below was carried out.

### 2.1 Methods and Materials

#### 2.1.1. The iTVR corpus

We first proceeded to create a corpus of videos for the analysis. We decided to search for videos of journalists reporting on the flooding that took place in Liguria (Italy) in October and in November 2014 in the Italian news portals Skytg24, Rainews24 and RaiTv.

The following criteria were followed to include videos in the corpus:

- The reporter was well visible (at least the arms, hands and face);
- The video was of relatively good quality;
- The audio was of relatively good quality;
- The communicative situation was homogeneous (reporter talking to the camera, no interaction with people; scenery in the background);
- The reporter was describing and/or talking about a natural event, like a disaster;
- The reporter was in sight for at least 10 seconds without interruptions (e.g., there were no interruptions such as footage showing the scenery that was being referred to);
- There were at least three videos of each reporter reporting the same event in different moments.

38 TV-news reports of variable duration by 9 Italian journalists (4 men and 5 women) were chosen for inclusion in the corpus. The videos were captured from full screen streaming using the Camtasia Studio 8 screen recorder software, which also records the system audio directly from sound card preserving the original audio quality. This corpus will be referred to as iTVR (Italian TV Reports corpus).

The multimodal annotation software ELAN (Wittenburg et al., 2006) was used to analyse the videos.

### 2.2 Analysis

A detailed qualitative analysis of the reporters' body language was carried out. The analysis focussed on the reporters' hand and arm gestures, though a note was made of the reporters' gaze, posture and body movements when these were conspicuous.

On the basis of this inspection, we defined three categories of speakers' dynamism, aimed at reflecting the extent to which the speakers were moving during their reports:

'Level 1'. The reporter is relatively still, and she hardly moves her arms and hands. When she wants to point at something, she does it at most with a movement of her eyes and gaze.

'Level 2'. The reporter is relatively still, but she is moving one arm and hand to emphasize a part of her report or a particular point in the scenery to which she is drawing the audience's attention.

'Level 3'. The reporter is moving constantly, and she turns a part of or the whole body together with her arm and hand to emphasize her discourse or to point to particular points in the scenery.

### 2.3 Results

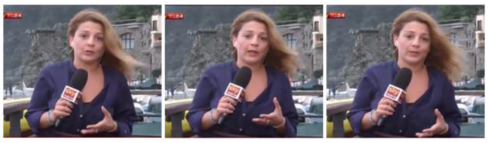
#### 2.3.1. Identification of Frequent Gestures in Emergency Communication

The reporters in the videos were assigned to the three levels of dynamism on the basis of the differences in the reporters' extent of gesturing

We identified two classes of arm and hand gestures that are distinctive of the different levels of dynamism.

The first is beats, that is, the rhythmic beating of a finger, hand or arm to accompany speech. Typically, beats involve up-and-down or back-and-forth hand movements that coincide with spoken clauses, breaks, or sentence ends (fig.1). Beats are not present in Level 1 videos, instead they are rather distinctive of Level 2 and 3, where the reporter emphasises words, sentences or speech rhythm in general.

Figure 1: Beat gestures



The second type of gestures that occur frequently in the corpus is pointing gestures (or deictics), that is, gestures that the journalists use to point to some place that they are referring to. This pointing gesture may be done with different degrees of 'extensiveness': from pointing that is merely hinted at with the gaze; to pointing that is done with the arm and hand while the journalists' body continues to stand still and face the camera; and finally to pointing that makes the journalists turn away from the camera and towards the place/situation that they are describing (fig. 2).

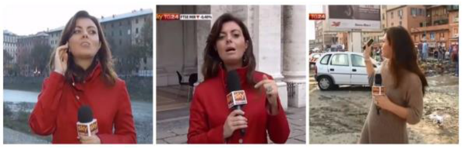


Figure 2: Pointing modes

These different extents in degree of movement correspond to the three identified levels of dynamism.

### 2.4 Discussion

Reporters appear to alternate between moments in which they gesture less (level 1) to moments when they gesture more (levels 2 and 3). Two types of gestures occur most frequently in the reporting of disasters. These are beats and deictics. When reporters use fewer gestures, they make little to no use of beats and deictics; when they gesture more they make movements that involve the whole body and extensive use of beats and deictics.

The frequent use of beats and deictics in this type of communication can be explained.

Beats are frequent in politicians' speeches (McNeill, 1992), especially when they have a cohesion function (when they serve to mark different points which are supposed to be crucial and coherent). Also, beats can appear to mark the word or phrase which introduces new characters, summarises the action, introduces new themes, etc. Neither politicians' speeches nor news reports involve a dialogical exchange or a real interaction with the interlocutor; the speakers - especially in tv reports - cannot count on back-channel feedbacks from the addressee and they probably need to mark the structure of the speech for clarity. Also, it has been shown that radio broadcasting news, for instance, can be characterised by "circumflex" intonation (a regularity in the use of pitch contours) and a constant and regular emphatic stress on words and syllables (Rodero, 2013). This typical rhythm that people commonly associate with news reading may also enhance the occurrence of beats, whose main function is rhythm beating.

As for deictics, it is quite natural for journalists reporting from a disaster site to point to the places and situations they are referring to, whether for clarity or to suggest to the cameraman where to shot. Also, their speech often presents adverbs referring to places and locations (here, there, etc.) that are naturally accompanied by pointings.

Research has shown that gestures and speech are interconnected (e.g., Goldin Meadow, 2005; Kendon, 1980). According to McNeill (1992), gestures and speech are synchronous at the semantic level, as they are co-expressive of the same underlying meaning, at the pragmatic level, as they co-occur to express the same pragmatic function; and at the phonological level, as gestures are temporally coordinated with the phonology of the utterances. Gestures and speech may also be constrained by the same contextual factors, accounting for individual differences, speakers' emotional involvement, etc. This is, however, still largely unexplored. Finally, the use of gestures during speech (co-speech gestures) is largely unconscious (Cienki & Müller, 2008). This makes body movements a way to explore underlying thoughts and emotive states of the speaker.

It is thus possible that when reporters' are reporting on disasters, their gesturing reflects their emotional involvement in the situation. It is also possible that the reporters' involvement will also show in some of the characteristics of the reporters' voices, such as pitch and intensity. The investigation of these aspects is the object of our next study.

## 3. Relating Reporters' Speech and Gestures: A Pilot Experiment

As discussed in the previous section, it is possible that reporters' use of gestures while communicating a disaster may betray different levels of emotional involvement. Specifically, increased gesturing may reflect an increased level of involvement in the situation, as an effect of their reduced control over their body language (reporters are likely to have learned to control their body language as part of their professional training).

Since gestures are synchronised with speech, it is likely that the increase in the reporters' gesturing may parallel an increased variability in the reporters' speech characteristics, and particularly those that are related to affect, such as pitch and intensity. These acoustic cues, in particular, are known to correlate with the emotional states of the speaker's voice (Juslin & Laukka, 2003).

To investigate these issues, this study addresses the relation of the reporters' differences in body dynamism to the acoustic correlates of pitch and intensity in the reporters' voices.

The aim was to test the following hypotheses:

H1: Variations in the reporters' body gesturing are paralleled by variations in pitch and intensity;

H2: The variation in the reporters' gestures, pitch and intensity can be perceived by the viewer; thus, this variation can be interpreted as a signal of emotional involvement;

H3: Extensive body gesturing is inversely related to perceived professionalism.

These hypotheses were tested in a two-part study. The first part is an acoustic analysis of the reporters' pitch and intensity variation patterns. These are then related to the differences in body dynamism that we observed in the qualitative analysis. The second part is an evaluation study testing the perception of the observed variation in the reporters' speech and gestures.

### 3.1 Procedure

#### 3.1.1. Materials

7 videos were selected randomly for each level of body dynamism (section 2.2) from the iTVR corpus. This created a corpus of 21 video samples. The duration of the videos was shortened to 10 seconds to include only the part that was most representative of the reporters' body dynamism and to allow the creation of an evaluation test

that was not too long.

### 3.1.2. Methods

The audio signal was extracted from the 21 videos. An acoustic analysis was carried out with Praat (Boersma, 2001). For both pitch and intensity, the listings of all the values were saved as a .txt file for further analysis.

For the evaluation test, three sets of stimuli were prepared. The first set used the 10-second videos but muted (Mute condition). The second set had audio tracks but no videos (Audio-only condition). The third set used the videos with the audios (Video condition).

In all the videos, the background around the reporters was removed with Adobe Premiere, so that the speakers appeared to speak on a black background (fig. 3). This was done to ensure that the background scenery (with views of the disaster) did not influence the participants' evaluation.

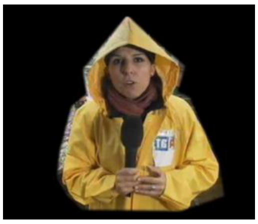


Figure 3: Example of a 10-second video with black background

Each set of stimuli contained a randomised list of the 21 items and was preceded by 4 videos selected additionally to create a trial session.

The test was administered online through Google Forms. The participants in the experiment were supposed to express - on a 1 to 5 point scale – their evaluation of: (a) the reporters' degree of involvement in the situation reported; and (b) the reporters' degree of professionalism. Forty evaluators completed the test.

## 3.2 Analysis

In the acoustic analysis the pitch and intensity listings were obtained for the 10-second audios with Praat.

The values were used to calculate the means and standard deviation (SD) of both pitch and intensity. For both acoustic cues, the SD was divided by the mean to normalize the data and remove inter-subjects differences, to obtain a measure called Pitch Variation Quotient (PVQ) (Hincks, 2004) and, by analogy, an Intensity Variation Quotient (IVQ).

The mean values of PVQ and IVQ were calculated for each level of dynamism, and the values for each level were compared to verify whether the reporters' variations in pitch and intensity parallel the reporters' variations in gesturing (H1).

For the evaluation study, all the scores obtained for each stimulus were averaged by level of dynamism. To verify the perception of the reporters' variation in speech and gestures, the mean scores were related to (1) the three levels of body dynamism that we had identified in the qualitative study and (2) the results of the acoustic analysis (H2 e H3).

At this stage of the investigation no statistical analyses have been carried out. These will be carried out on a larger data sample.

## 3.3 Results

Table 1 shows the Pitch Variation Quotient (PVQ) and the Intensity Variation Quotient (IVQ) in relation to the three levels of dynamism that were presented in Section 2.2

| Level | PVQ* | IVQ** |
|---|---|---|
| **1** | 0.16 | 0.10 |
| **2** | 0.24 | 0.10 |
| **3** | 0.22 | 0.08 |

* Pitch Variation Quotient; ** Intensity Variation quotient.
Level **1**= "Idle"; Level **2** = Beats only (still body); Level **3** = Beats, Pointings, Body moves.

Table 1: PVQ and IVQ in relation to the three levels of speakers' body dynamism.

The results confirm H1 only partially. For pitch, the values of PVQ show that there is an increase in overall variation from Level 1 to Level 2, while the values for Level 3 are slightly lower than those of Level 2. As for intensity, the data do not provide support to our hypothesis, and show equal intensity values for Level 1 and 2 and slightly lower for Level 3.

Table 2 shows the mean values of the evaluators' ratings of the three sets of stimuli. The data show some interesting trends.

| Level | M* | | A** | | V*** | |
|---|---|---|---|---|---|---|
| | **a** | **b** | **a** | **b** | **a** | **b** |
| **1** | 2.41 | 3.44 | 2.83 | 3.46 | 2.51 | 3.40 |
| **2** | 3.46 | 3.36 | 3.44 | 3.11 | 3.33 | 3.02 |
| **3** | 3.44 | 3.02 | 3.46 | 3.21 | 3.44 | 2.99 |

*Mute condition; ** Audio only condition; *** Video condition.
a= question about involvement; b= question about professionalism.

Table 2: Mean values of the evaluators' ratings of the three types of stimuli: Muted videos (M), Audio only (A), regular videos (V).

All reporters were considered to be less involved at Level 1 of body dynamism than in the other levels (Column a) in all conditions. In other words, reporters that move little were perceived as less involved than reporters that move more while reporting on emergency situations. This confirms H2. The fact that the tendency is true also in the audio only condition suggests that listeners are able to interpret speakers' voice qualities as involvement in the same way as they do with gestures and body movements. In fact, the data suggest that when no video is present, the audio data has a stronger effect than the video, at least at Level 1 of body dynamism. As far as professionalism is concerned (Column b) an opposite trend can be detected: the more a reporter moves the less he/she is perceived as professional – as shown by the values decreasing from Level of body dynamism 3 to level 1. This confirms H3.

### 3.4 Discussion

We predicted that variation in both pitch and gestures increases as speakers increase their involvement in the reported event. The results of our study provide only partial support to our hypothesis. However, it should be noted that the audio data extracted from the videos were rather noisy, due to the fact that the speakers were reporting from emergency scenes, and so the background noise may have affected the results of the acoustic analysis.

One of the aims of our evaluation experiment was to test whether greater gesturing is perceived as greater involvement. The results confirm our hypothesis and show that, when reporters gesture more or more extensively (see section 2.2), they are perceived as more emotionally involved in the situation they are reporting. Also, there is an inverse relation between perceived involvement, as is reflected by gesturing, and perceived professionalism: the more the reporters in the videos were using gestures the less professional they were judged. This result confirms our hypothesis that extensive body gesturing is inversely related to perceived professionalism.

## 4. Towards an ontology of gestures used in the communication of emergencies

As reviewed in the introduction, research on non-verbal language can be used to extract data on speakers' emotional involvement. In this work we identified two gestures that seem to be characteristic of emotional reporting in emergency communication. In this section the ontology of these two gestures is proposed.

### 4.1. Ontology of Beats and Pointing Gestures

Our analysis showed that the two gestures that are most commonly used in emergency communication are beats and deictics. To develop the ontology for these gestures we drew on influential nonverbal classification schemes (Efron, 1941; Ekman and Friesen, 1969; McNeill 1992; Kendon 2004). The ontology was then refined by reference to other notation systems and coding schemes that were recently developed in association with automatic human gesture recognition and synthesis (e.g., Bressem, 2008; Kipp, Neff & Albrecht, 2007).

The ontology that was developed for beat gestures consists of a decision tree (Figure 4) that can be considered a 'filter' for defining the gesture.

A beat gesture has to satisfy some necessary hierarchical conditions to be defined as a beat gesture:

- It has to be a repetitive movement that follows the speaker's speech rhythm. This movement has a certain frequency and duration that can be measured;
- It has to be a straight movement. It can be an up-down or a back-forth movement;
- The movement has to show a certain muscle tension. High-level labels (e.g., flat hand/spread or single finger/bent) represent hand shapes and orientation and can be assigned to any gesture (Figure 5).

A decision tree of the same model was developed for pointing gestures. We took into consideration the most prototypical pointing gesture, that is the one performed to with the hand or the arm, excluding other pointing strategies (e.g., gaze). The schema lists all the necessary conditions that must be satisfied to assign the pointing label to a gesture.

The following conditions have to be hierarchically satisfied:

- The gesture needs to trace a well-defined path. It means that it needs to be a movement that follows a clear direction, i.e., moves clearly towards something;
- The final part of the movement is usually linear (not circular or spiral);
- The gesture is usually held for a while at its furthest extent;
- Hands must have a certain shape or a certain muscle tension. High-level labels (e.g., index finger - palm down or index finger - palm up) represent hand shapes and orientation and can be given to any gesture.

## 5. Conclusion

This study aimed to examine whether speech and non-verbal language can be used to detect the speakers' emotional involvement. For this purpose, it analysed a corpus of videos on journalists reporting on natural disasters.

The study shows that, in their emergency reports, journalists vary their voice characteristics and gestures considerably. In particular, their voice may span from lower to higher levels of pitch and intensity; their bodies also show different levels of dynamism. Two gestures seem to be characteristic of this kind of reporting: beats and deictics. The results of an evaluation test showed that the reporters' variations in voice pitch and gesturing are perceived as a signal of emotional involvement and affect the perception of the reporters' professional image.

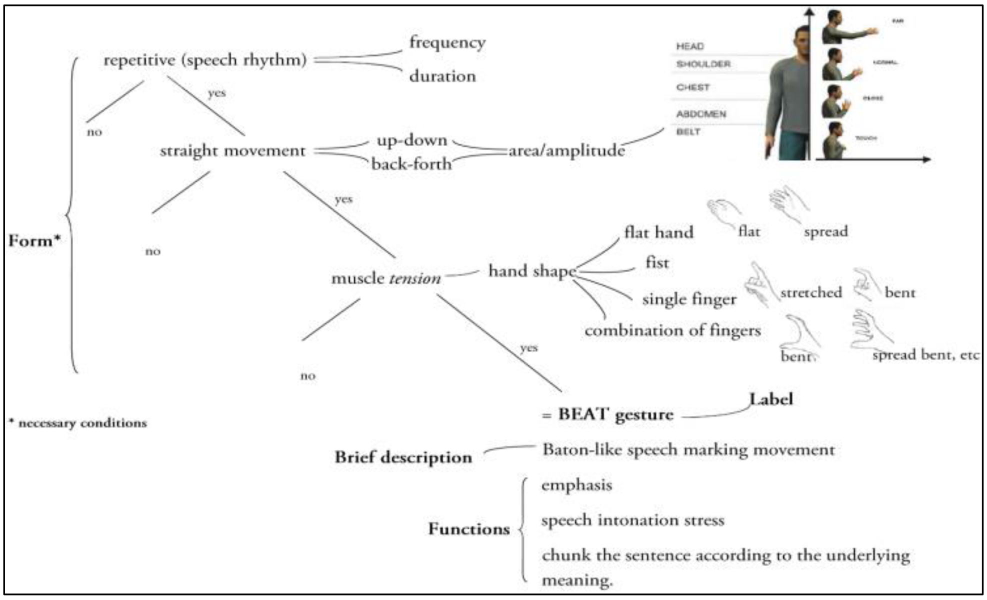


Figure 4: Beat gestures decision tree. Hands and body illustrations from Bressem (2008) and Kipp et al. (2007).

Our results should be considered preliminary, and more work will done to expand this research. We are planning to gather quantitative data on a wider repository of gestures (head, gaze, and other categories of hand gestures) and extract information about gesture occurrences, duration, latency, etc. Also, we will carry out additional acoustic analyses, including measurements of speech rate and vocal perturbation. With more data we will also run statistical analyses and investigate the correlation between gestures and voice cues.

Though preliminary, the results of this study show the importance of studying emotional involvement that can be expressed *beyond* the speaker's words.

Reporters play a key role in the delivery of information. In emergencies, reporters' voice or body language may reflect feelings of anxiety or fear that are not conveyed by the words alone. This may impact on the way the message is received by the public, with consequences on their actions or thoughts. Thus, this research can provide information that is useful for training end-users and spokespeople in emergency communication.

Finally, the evidence that speakers' modify their voice and

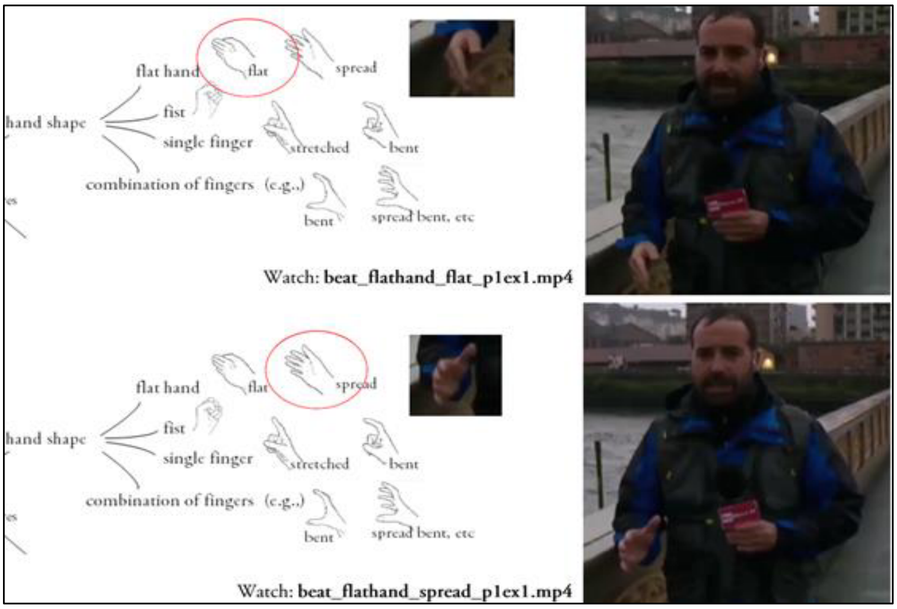


Figure 5: Beat Gestures. Hand shapes and gestures areas. Hands illustrations from Bressem (2008).

body language patterns as an effect or a reaction to a disaster can be used to obtain indirect information about the gravity of the disaster. These data can be used for improving the communication protocols in an emergency management system.

## 6. References

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Bressem, J. (2008). Notating gestures—Proposal for a form based notation system of coverbal gestures. *Unpublished manuscript. Retrieved from http://www. janabressem. de/publications. html*.

Cienki, A., & Müller, C. (2008). *Metaphor and gesture.* John Benjamins Publishing.

de Gelder, B. (2009). Why bodies? twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences, 364*(1535), 3475-3484. doi:10.1098/rstb.2009.0190 [doi]

Efron, D. (1941). Gesture and environment.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*(1), 49-98.

Ekman, P., & Friesen, W. V. (1977). Facial action coding system.

Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings* Elsevier.

Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think* Harvard University Press.

Hincks, R. (2004). Processing the prosody of oral presentations. *InSTIL/ICALL Symposium 2004,*

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. *The New Handbook of Methods in Nonverbal Behavior Research,* , 65-135.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication, 25*, 207-227.

Kendon, A. (2004). *Gesture: Visible action as utterance* Cambridge University Press.

Kipp, M., Neff, M., & Albrecht, I. (2007). An annotation scheme for conversational gestures: How to economically capture timing and form. *Language Resources and Evaluation, 41*(3-4), 325-339.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* University of Chicago press.

Mehrabian, A. (1972). *Nonverbal communication* Transaction Publishers.

Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*(3), 248.

Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology, 6*(1), 109.

Rodero, E. (2015). The principle of distinctive and contrastive coherence of prosody in radio news: An analysis of perception and recognition. *Journal of Nonverbal Behavior, 39*(1), 79-92.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1), 227-256.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. *Proceedings of LREC, 2006* 5th.