





**Multimodal System Input**  
**Professor Doctor Sebastian Moller**  
**Quality and Usability Lab**  
**Technische Universität Berlin**  
**Emotion Recognition**

(Refer Slide Time: 00:18)



**Multimodal System Input: Emotion recognition.**  
Prof. Dr.-Ing. Sebastian Möller  
Quality and Usability Lab  
Technische Universität Berlin



Not only text or linguistic information can be extracted from auditory or visual signals but also the emotion of the user can be extracted from that. And this is the task of emotion recognition.

(Refer Slide Time: 00:31)

**Emotion recognition.**

**Basic emotions** (from Ekman & Friesen, 1984; Scherer, 2003):

- Boredom / surprise
- Fear
- Disgust
- Anger
- Joy
- Sadness

**Relevant channels to transmit emotions:**

- Visual (facial expression)
- Auditory (prosody)
- Physiological (mostly been disregarded, but become more important)



Actually there exist several classifications of emotions.

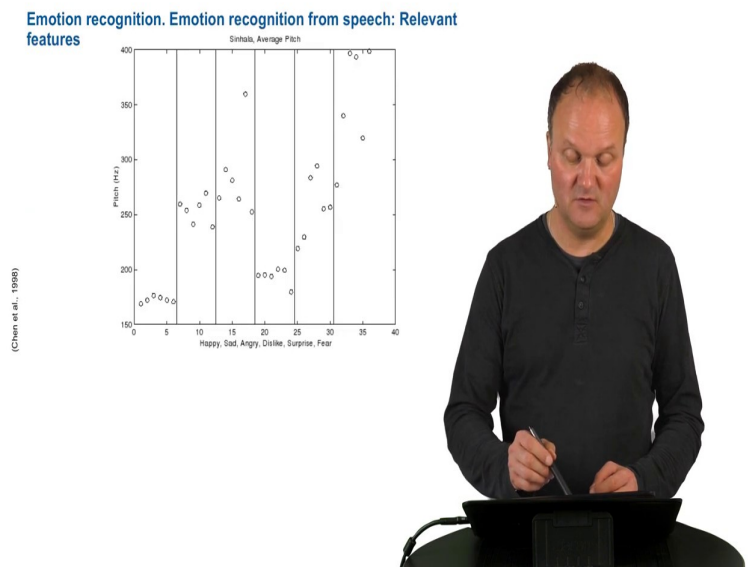
The most basic one is the one from Ekman and Friesen and Scherer which differentiate between boredom or surprise, fear, disgust, anger, joy and sadness as basic emotions which humans are able to show.

These emotions are transmitted through different channels, mainly the visual channel, facial expression, the auditory channel mainly including the prosody that is the movement of the pitch and the energy and the duration of the speech signals.

And they also showed in the physiological way which is mostly has been disregarded or has been disregarded in the past but becomes more and more important with availability of sensors in order to sense these physiological signals.

We will now in the following concentrate in the visual and the auditory features which can be used

(Refer Slide Time: 01:28)

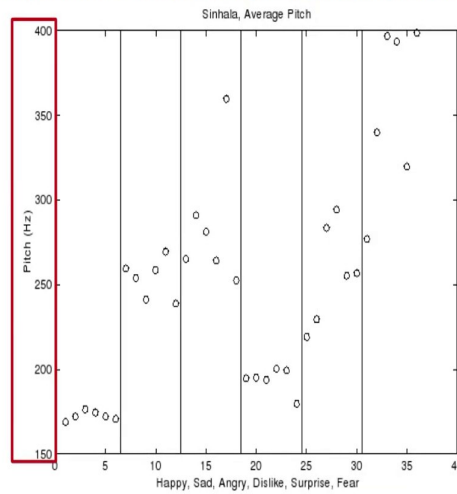


in order to recognize the emotions. Here we see an example.

(Refer Slide Time: 01:33)

### Emotion recognition. Emotion recognition from speech: Relevant features

(Chen et al., 1998)

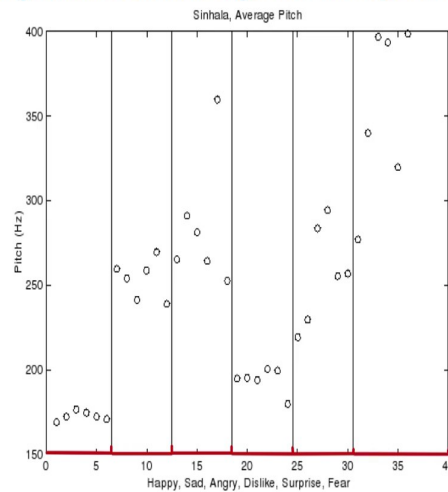


The pitch frequency that is the fundamental frequency or the perception of fundamental frequency of the voice for the different classes of emotions extracted from a corpus of relevant speech and separated into happy, sad, angry, dislike, surprise and fear emotions.

(Refer Slide Time: 01:54)

### Emotion recognition. Emotion recognition from speech: Relevant features

(Chen et al., 1998)

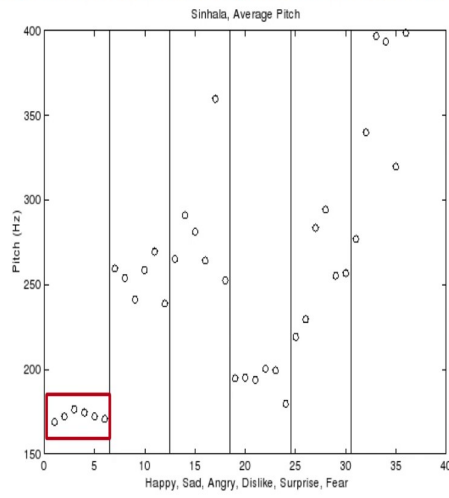


And you see that the pitch alone helps a little bit to classify for example happy emotions, so this lower pitch here,

(Refer Slide Time: 02:02)

### Emotion recognition. Emotion recognition from speech: Relevant features

(Chen et al., 1998)

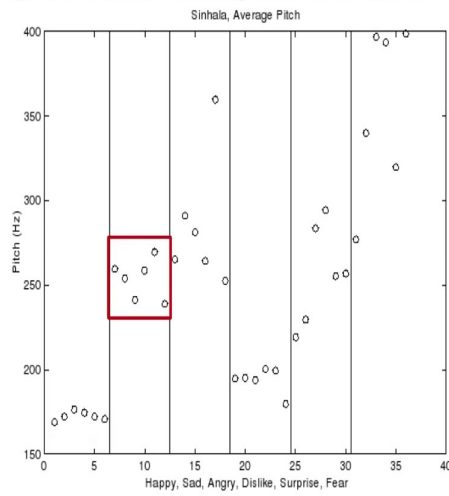


sad emotions

(Refer Slide Time: 02:03)

### Emotion recognition. Emotion recognition from speech: Relevant features

(Chen et al., 1998)

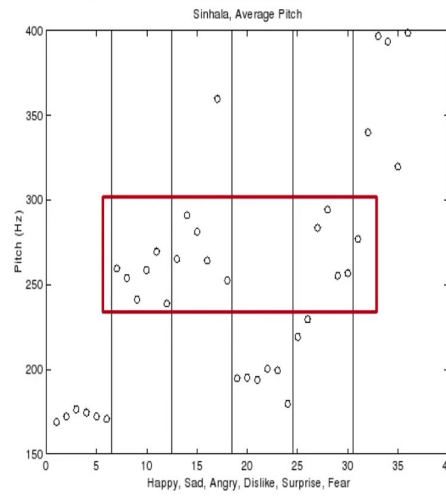


so this higher pitch here but there is lot of overlap between

(Refer Slide Time: 02:07)

### Emotion recognition. Emotion recognition from speech: Relevant features

(Chen et al., 1998)



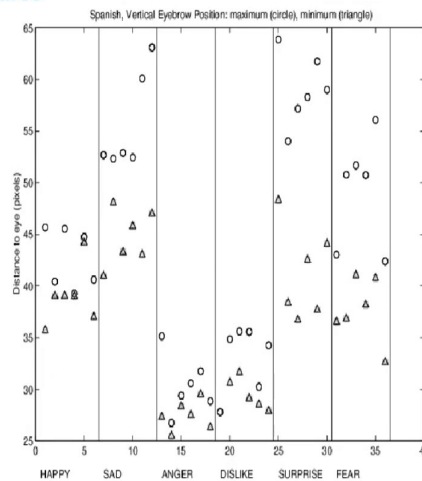
the classes.

So you can imagine that the pitch frequency alone is not sufficient for exactly, undoubtedly classify emotions. You can take other information, for example visual information

(Refer Slide Time: 02:22)

### Emotion recognition. Emotion recognition from facial expression: Relevant features

(Chen et al., 1998)

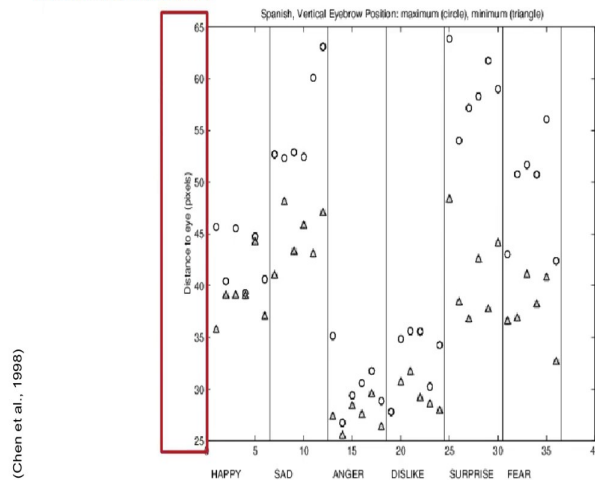


as we see here.

Here we have the distance

(Refer Slide Time: 02:26)

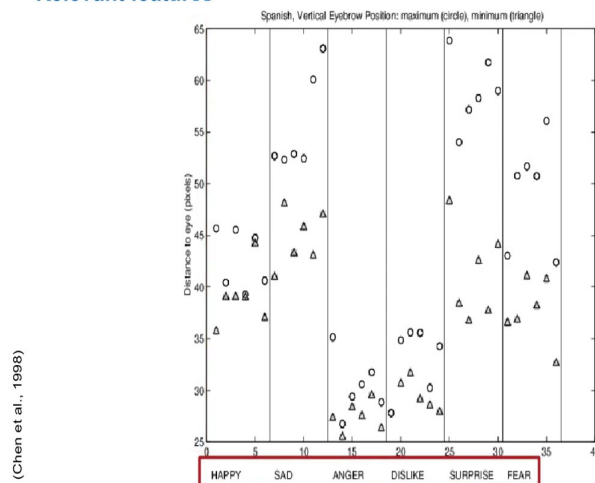
### Emotion recognition. Emotion recognition from facial expression: Relevant features



to the eyes and we have once again the 6 basic emotions here

(Refer Slide Time: 02:31)

### Emotion recognition. Emotion recognition from facial expression: Relevant features



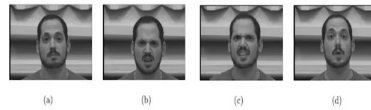
and you see that some of them can be differentiated on the basis of this parameter and some others cannot.

So apparently it is good to combine audio and visual information. And it seems that audio and visual information is not equally relevant for all types of emotions.

(Refer Slide Time: 02:51)

Emotion recognition. Audio-visual emotion recognition:

Experiments by De Silva et al.



- **Results:**
  - **Video-dominated:** Anger, Happiness, Surprise
  - **Audio-dominated:** Sadness, Fear
  - **Mixed:** Dislike

(De Silva et al., 1997)



Actually results from the experiments by De Silva et al showed that they are video-dominated emotions like anger, happiness and surprise.

And there are others which you can better extract from the audio signal, like sadness and fear and there are some which are illustrated in both types of signals like dislike.

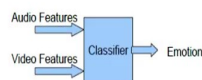
So apparently it is good to combine video information with audio information in order to classify speech, classify emotions.

(Refer Slide Time: 03:23)

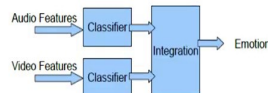
Emotion recognition. Audio-visual emotion recognition: Fusion

- Once again two possibilities:

- **Feature fusion**



- **Decision fusion**



The way this can be handled is the same as we have seen for audiovisual speech recognition.

So we can perform a feature fusion by connecting the audio features to the video features and then using one classifier only, all the decision fusion or probability fusion where we take individual classifiers for the audio features first, for the video features and then integrate that information into one recognized emotion in the end.

(Refer Slide Time: 03:55)

### References.

- Chen, L.S., Huang, T.S., Miyasato, T., Nakatsu, R. (1998). Multimodal Human Emotion/Expression Recognition, Proceedings of the 3rd. International Conference on Face & Gesture Recognition, p.366, April 14-16.
- De Silva, L.C., Miyasato, T., Nakatsu, R. (1997). Facial Emotion Recognition Using Multimodal Information, Proceedings of 1997 International Conference on Information, Communications and Signal Processing (ICICS), Singapore, pp. 397-401.
- Ekman, P., Friesen, W.V. (1984). *Unmasking the Face*. Consulting Psychologists Press, Palo Alto CA.
- Scherer, K., Ekman, P. (1982). *Handbook of Methods in Nonverbal Behavior Research*. Cambridge University Press.
- Scherer, K., Johnstone, T., Klasmeyer, G., (2003). Vocal expression of emotion, in: *Handbook of the Affective Sciences*, R. Davidson, K. Scherer, H. Goldsmith (Eds.), New York: Oxford University Press, 433—456.