



Feature selection in acted speech for  
the creation of an emotion recognition  
personalization service

*C. Anagnostopoulos*

*University of the Aegean  
Cultural Technology and Communication Dpt.*

# What emotion is?

---

□ ....“Everyone knows what an emotion is, until asked to give a definition”....

- Beverly Fehr and James Russell -

□ Emotions are important for:

- motivation, perception, ability, creativity, attention, organization, learning, memory and decision ability

# Problem definition

---

Building emotion recognition modules is important for improving human-computer interaction or building personalisation service.

What happens during people communication as emotional beings?

New branch of Artificial Intelligence

## Emotional Intelligence

- ✓ Development of more user friendly GUIs
- ✓ Acquire information during interaction

# Emotional Intelligence

---

The field of emotional intelligence is dedicated to making technology more emotionally engaging and responsive, and inherently more naturally useable.

System that facilitates natural interaction in addition to providing heightened realism and aesthetic desirability

E.g. Stress levels in a driver.

# Computer emotion recognition

---

ER refers to the identification of the emotional state of a user that interacts with a system.

1. Biosensors

2. Face recognition modules

3. Recognition from speech (w/o linguistic information)

4. Various combinations of the above

# Biosensors

---

Different emotional expressions produce different changes in activity:

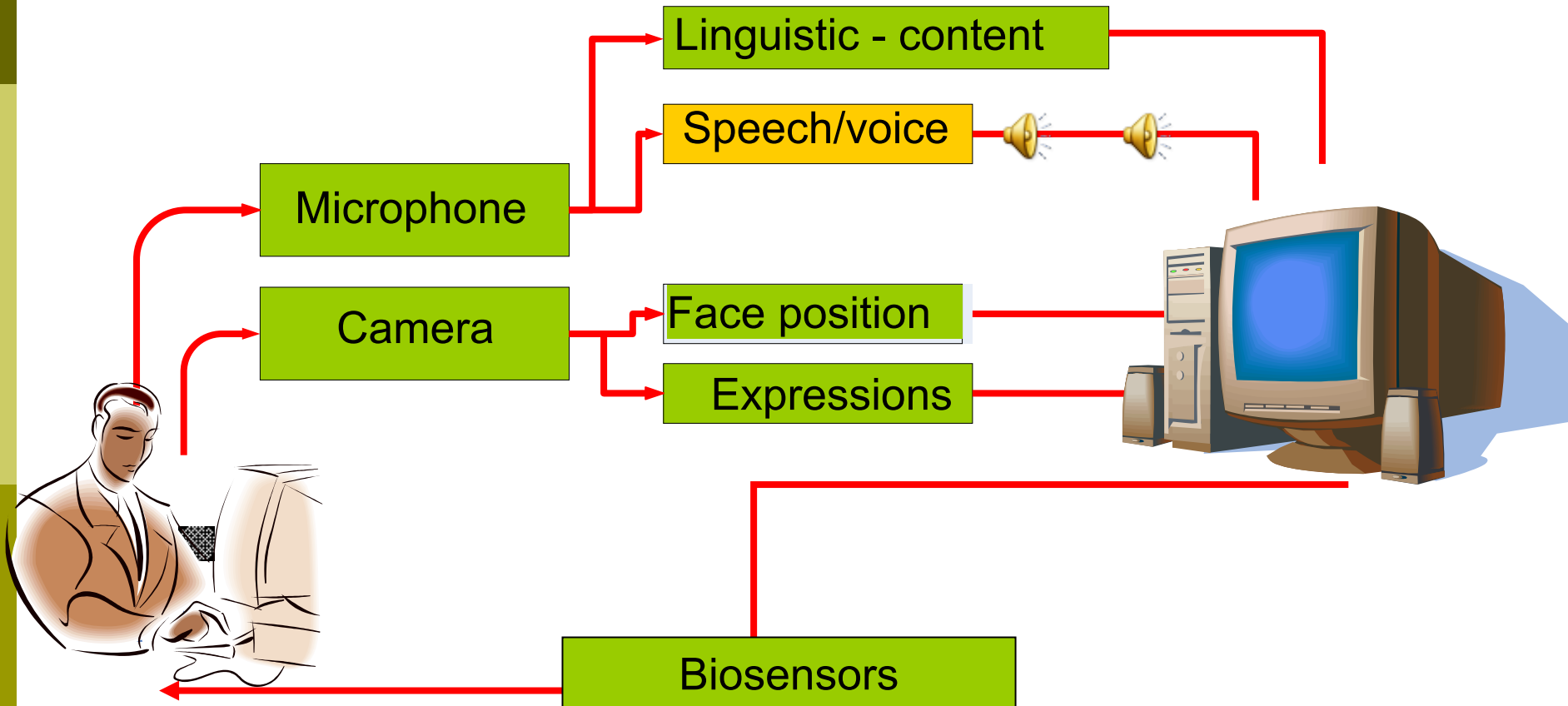
Anger: increased HR and skin temperature

Fear: increased HR, decreased skin temperature

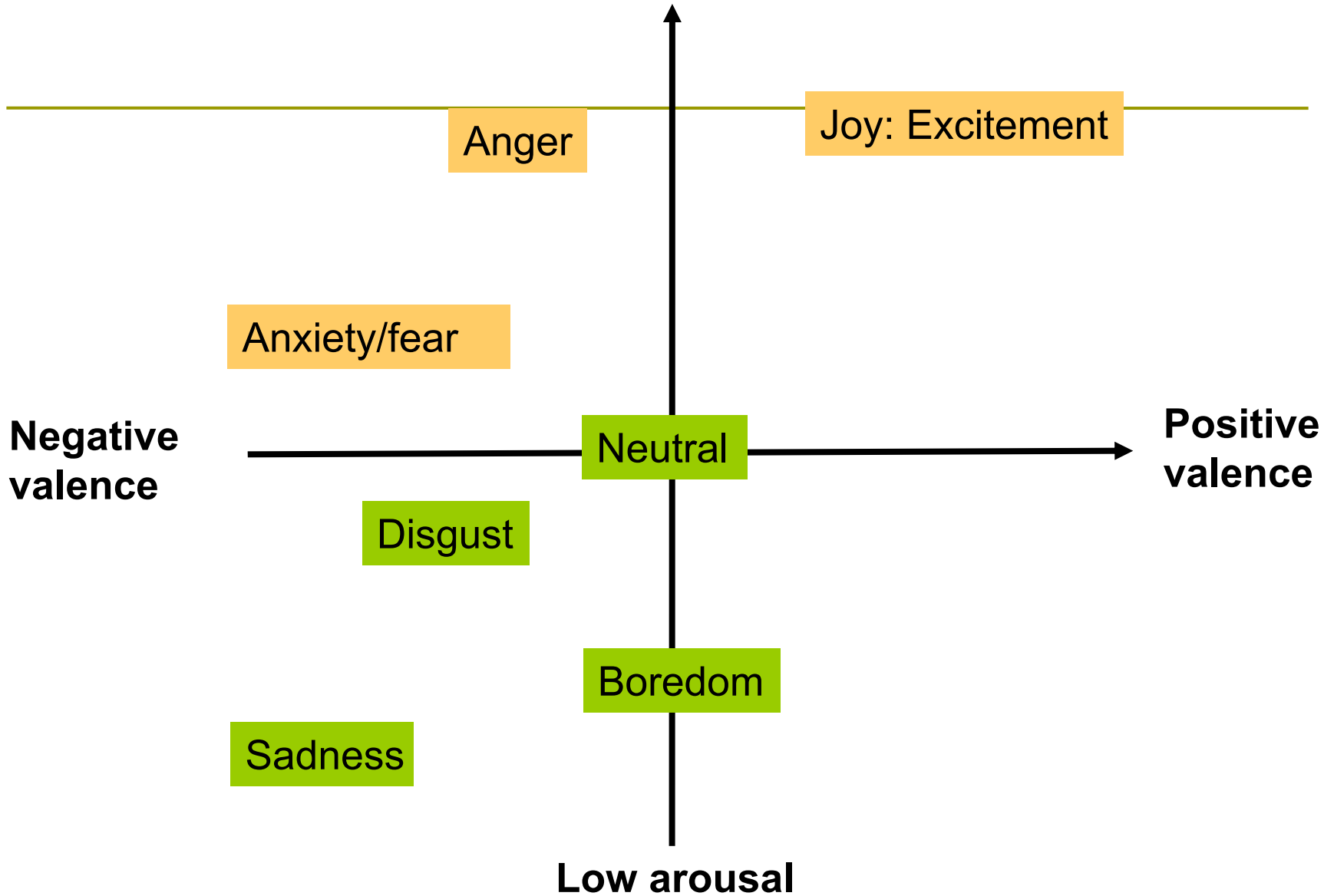
Happiness: decreased HR, no change in ST

Easily integrated with external channels (face and speech)

# Emotion recognition system



High arousal



Anger

Joy: Excitement

Anxiety/fear

Disgust

Boredom

Sadness

# Data based used

---

We used the open access Berlin Database

<http://pascal.kgw.tu-berlin.de/emodb/>

Includes: 535 speech recordings

➤ 10 actors: 5 male, 5 female

➤ 10 utterances (in German)

➤ 7 emotions: **Anger, joy, anxiety**

neutral, boredom, disgust and sadness

# Berlin database

---

With Berlin Database:

Gender recognition experiments

Emotion recognition experiments

Speaker dependent – independent

Utterance independent - independent

This work focuses in the speaker and utterance (phrase) independent framework.

# Berlin database

---

An example

Der Lappen liegt auf dem Eisschrank

Neutral	
Anger	
Anxiety/fear	
Joy/Excitement	
Boredom	

# Basic scope

---

Identify the sound features/measurements that provide important information concerning the emotional state.

Information of the linguistic content is not used.

# Sound measurements

---

- Fundamental frequency (F0 or pitch)
  - Energy
  - Formants (F1, F2, F3, F4, F5)
  - 12 Mel- Frequency Cepstral Coefficients (MFCCs)
- 
- Some others: periodicity, rhythm, voiced/unvoiced time ratio

# Sound measurements

---

All 19 features are measured in short time periods.

At the end of each phrase we calculate 7 measurements:

the mean, the standard deviation, the minimum value, the maximum value, the range (max-min) of the original contour and the mean and standard deviation of the contour gradient.

# Sound measurements (choice)

---

133 features -> Reduce through WEKA

Feature evaluator: CfsSubSetEval

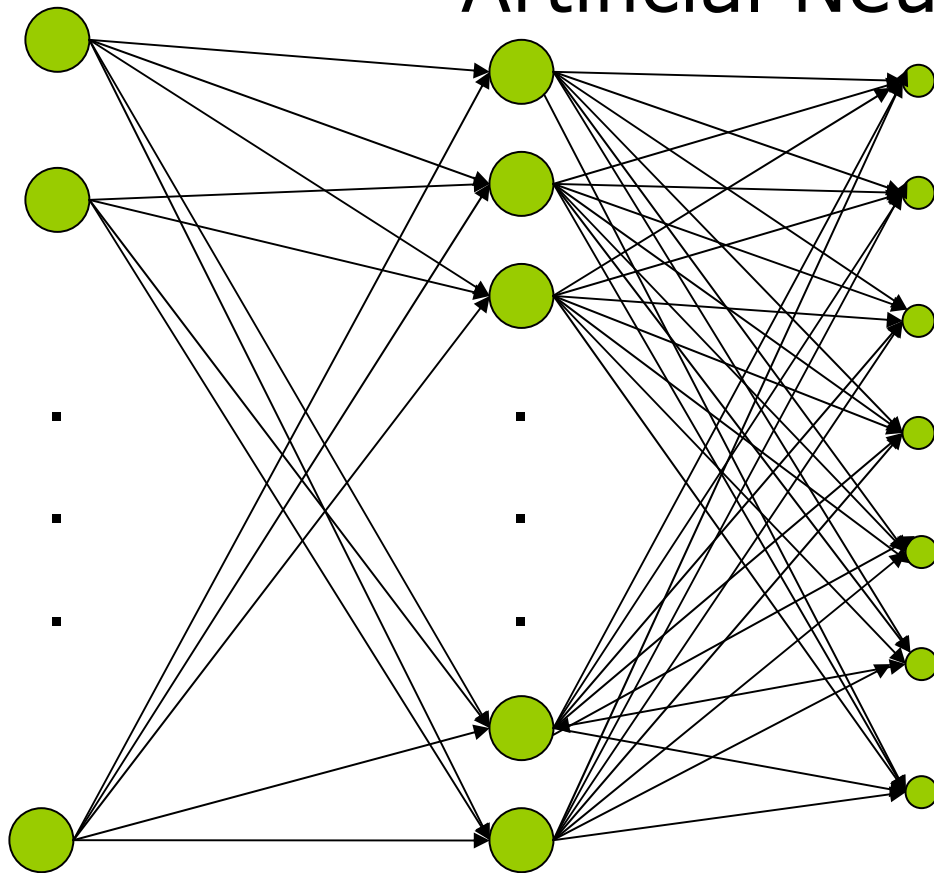
Feature search method: BestFirst

Finally 23 features have been selected (you can see them in the next slide)

<b>Prosodic Feature</b>	<b>Mean</b>	<b>Std</b>	<b>Mean of derivative</b>	<b>Std of derivative</b>	<b>Max</b>	<b>Min</b>	<b>Range</b>
<b>Pitch</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>MFCC1</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>
<b>MFCC2</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>	<b>21</b>
<b>MFCC3</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>
<b>MFCC4</b>	<b>29</b>	<b>30</b>	<b>31</b>	<b>32</b>	<b>33</b>	<b>34</b>	<b>35</b>
<b>MFCC5</b>	<b>36</b>	<b>37</b>	<b>38</b>	<b>39</b>	<b>40</b>	<b>41</b>	<b>42</b>
<b>MFCC6</b>	<b>43</b>	<b>44</b>	<b>45</b>	<b>46</b>	<b>47</b>	<b>48</b>	<b>49</b>
<b>MFCC7</b>	<b>50</b>	<b>51</b>	<b>52</b>	<b>53</b>	<b>54</b>	<b>55</b>	<b>56</b>
<b>MFCC8</b>	<b>57</b>	<b>58</b>	<b>59</b>	<b>60</b>	<b>61</b>	<b>62</b>	<b>63</b>
<b>MFCC9</b>	<b>64</b>	<b>65</b>	<b>66</b>	<b>67</b>	<b>68</b>	<b>69</b>	<b>70</b>
<b>MFCC10</b>	<b>71</b>	<b>72</b>	<b>73</b>	<b>74</b>	<b>75</b>	<b>76</b>	<b>77</b>
<b>MFCC11</b>	<b>78</b>	<b>79</b>	<b>80</b>	<b>81</b>	<b>82</b>	<b>83</b>	<b>84</b>
<b>MFCC12</b>	<b>85</b>	<b>86</b>	<b>87</b>	<b>88</b>	<b>89</b>	<b>90</b>	<b>91</b>
<b>Energy</b>	<b>92</b>	<b>93</b>	<b>94</b>	<b>95</b>	<b>96</b>	<b>97</b>	<b>98</b>
<b>F1</b>	<b>99</b>	<b>100</b>	<b>101</b>	<b>102</b>	<b>103</b>	<b>104</b>	<b>105</b>
<b>F2</b>	<b>106</b>	<b>107</b>	<b>108</b>	<b>109</b>	<b>110</b>	<b>111</b>	<b>112</b>
<b>F3</b>	<b>113</b>	<b>114</b>	<b>115</b>	<b>116</b>	<b>117</b>	<b>118</b>	<b>119</b>
<b>F4</b>	<b>120</b>	<b>121</b>	<b>122</b>	<b>123</b>	<b>124</b>	<b>125</b>	<b>126</b>
<b>F5</b>	<b>127</b>	<b>128</b>	<b>129</b>	<b>130</b>	<b>131</b>	<b>132</b>	<b>133</b>

# Classifier

Artificial Neural Network: (MLP)



23-30-7

# 5 experiments

---

Experiment	Testing set Speaker code (gender)	Training set Speaker code (gender)
1	10,11,12,15 (male), 09,13,14,16 (female)	03 (male), 08 (female)
2	03,11,12,15 (male), 08,13,14,16 (female)	10 (male), 09 (female)
3	03,10,12,15 (male), 08,09,14,16 (female)	11 (male), 13 (female)
4	03,10,11,15 (male), 08,09,13,16 (female)	12 (male), 14 (female)
5	03,10,11,12 (male), 08,09,13,14 (female)	15 (male), 16 (female)

# Overall results - 7 classes

	High arousal emotions			Low arousal emotions			
	anger	happiness	anxiety/ fear	boredom	disgust	sadness	neutral
Anger	76 (60.3%)	23 (18.3%)	18 (14.3%)	0 (0.0%)	8 (6.3%)	0 (0.0%)	1 (0.8%)
Happiness	22 (31.0%)	30 (42.3%)	6 (8.5%)	1 (1.4%)	9 (12.7%)	0 (0.0%)	3 (4.2%)
anxiety /fear	28 (40.6%)	6 (8.7%)	27 (39.1%)	6 (8.7%)	1 (1.4%)	1 (1.4%)	0 (0.0%)
Boredom	1 (1.2%)	2 (2.5%)	12 (14.8%)	43 (53.1%)	3 (3.7%)	9 (11.1%)	11 (13.6%)
Disgust	6 (13.0%)	9 (19.6%)	2 (4.3%)	11 (23.9%)	16 (34.8%)	0 (0.0%)	2 (4.3%)
Sadness	0 (0.0%)	2 (3.2%)	1 (1.6%)	11 (17.7%)	3 (4.8%)	37 (59.7%)	8 (12.9%)
neutral	1 (1.3%)	2 (2.5%)	3 (3.8%)	20 (25.0%)	0 (0.0%)	10 (12.5%)	44 (55.0%)

e.g. 126 phrases (emotion anger), 76 were correctly classified.

# Results - 2 hyperclasses

---

	High arousal	Low arousal
High arousal	236/266 (88.7%)	30/266 (11.3%)
Low arousal	41/269 (15.2%)	228/269 (84.8%)

# Literature

<b>Ref.</b>	<b>Features used</b>	<b>Emotions/classes</b>	<b>Overall performance</b>
[2]	Up to 26	8	80.0% (speaker dependent)
[5]	Up to 34	6	72.2%
[11]	30	7 (Berlin Database)	81.1% (speaker independent)
[12]	20	8	85.4% (speaker dependent)
[20]	20	7 (MPEG-4 + neutral)	71.6% (speaker independent)
[21]	3	4	75.7%
[22]	16	4	74.3%
This paper	23	7 (Berlin Database)	51% (Speaker and utterance independent)

# Conclusions

---

The researchers usually deal with elicited and acted emotions in a lab setting from few actors. However, in the real problem, different individuals reveal their emotions in a diverse degree and manner. There are also many differences between acted and spontaneous speech.

Concluding this work, we should emphasize the difficulty of the speech emotion recognition problem. In this interdisciplinary field of research, aspects of psychology and physiology are not always considered and literature still offers ideas rather than solutions.

After all, consider how difficult is for an individual to understand the emotional state of a person.