

Speech Personality

Computational Social Intelligence - Lecture 18

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

<http://www.dcs.gla.ac.uk/vincia>
Alessandro.Vinciarelli@glasgow.ac.uk



University
of Glasgow

EPSRC

Engineering and Physical Sciences
Research Council



Texts (see Moodle)

This lecture is based on the following text (available on Moodle):

- G.Mohammadi and A.Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic features", IEEE Transactions on Affective Computing 3(3): 273-284, 2012.

Outline

- Introduction
- Computational Paralinguistics
- Trait Prediction
- Conclusion

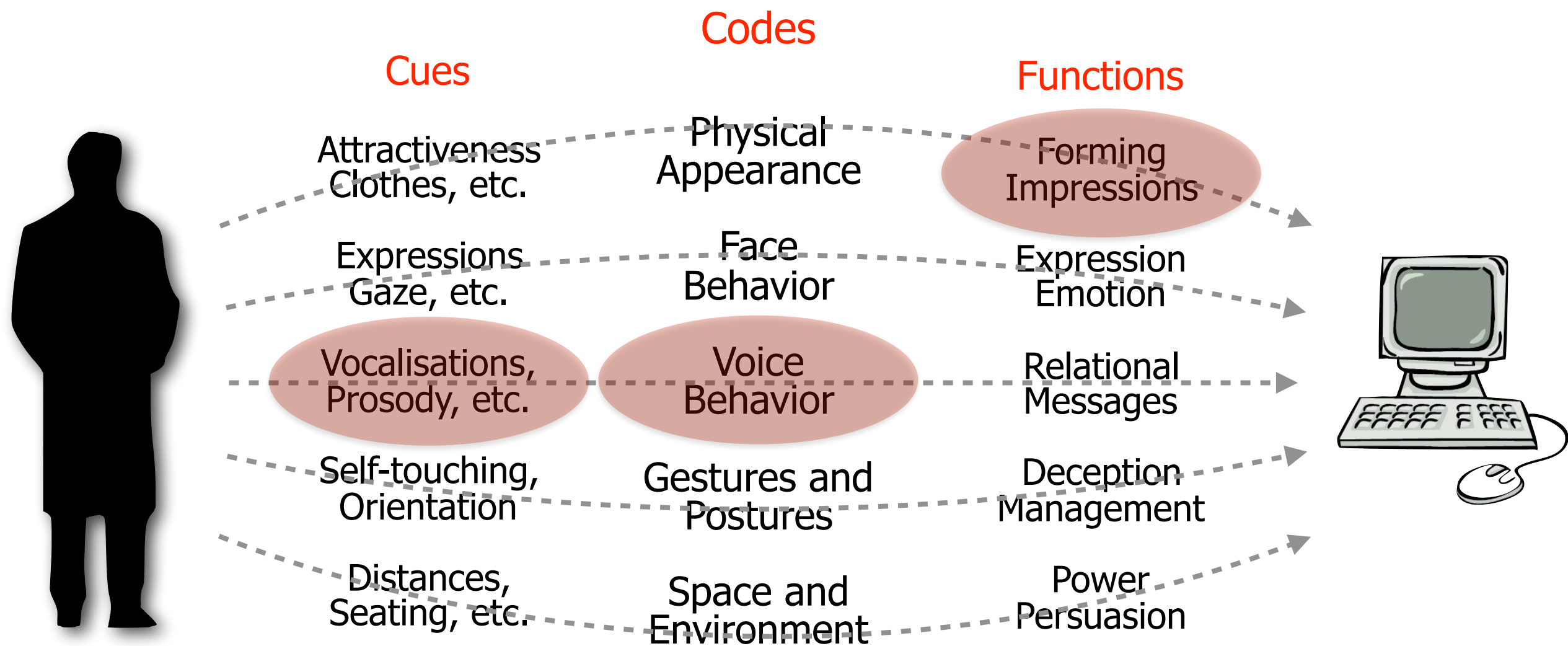
Outline

- Introduction
- Computational Paralinguistics
- Trait Prediction
- Conclusion

Speech and Personality

“[...] judgments made from speech alone rather consistently [have] the highest correlation with whole person judgments.”

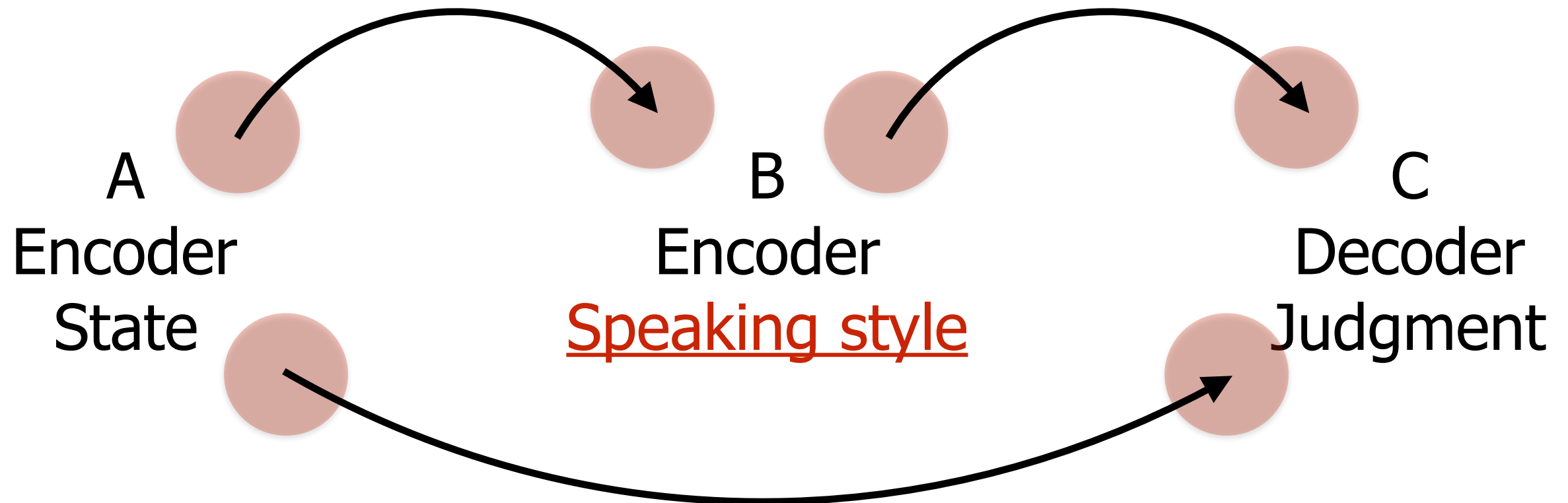
Ekman, Friesen, O'Sullivan, Scherer, "Relative importance of face, body, and speech in judgments of personality and affect", Journal of Personality and Social Psychology, 38(2):270-277, 1980.



Richmond and McCroskey, "Nonverbal Behaviors in Interpersonal Relations",
Allyn and Bacon, 1995

How does the encoder manifest her/his state through her/speaking style?

How do decoders interpret the speaking style of the encoder



What is the state the decoder attributes to the encoder?

Recap

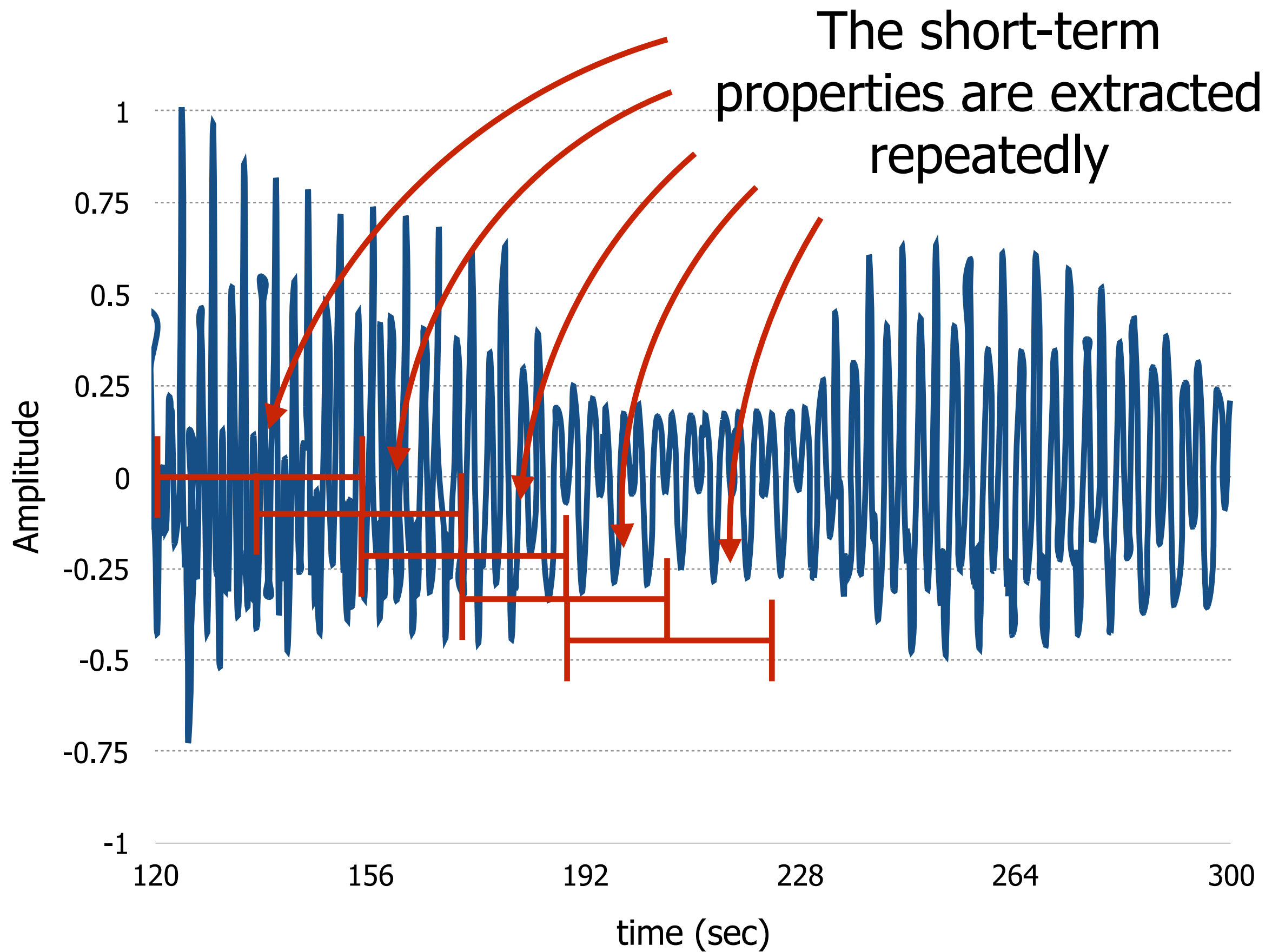
- The short-term properties of the speech signals provide information about the speaking style (how people speak);
- The speaking style influences the impression people develop about the speaker;
- It is possible to infer the personality traits people attribute to a speaker from her/his speaking style.

Outline

- Introduction
- **Computational Paralinguistics**
- Trait Prediction
- Conclusion

Short-Term Features

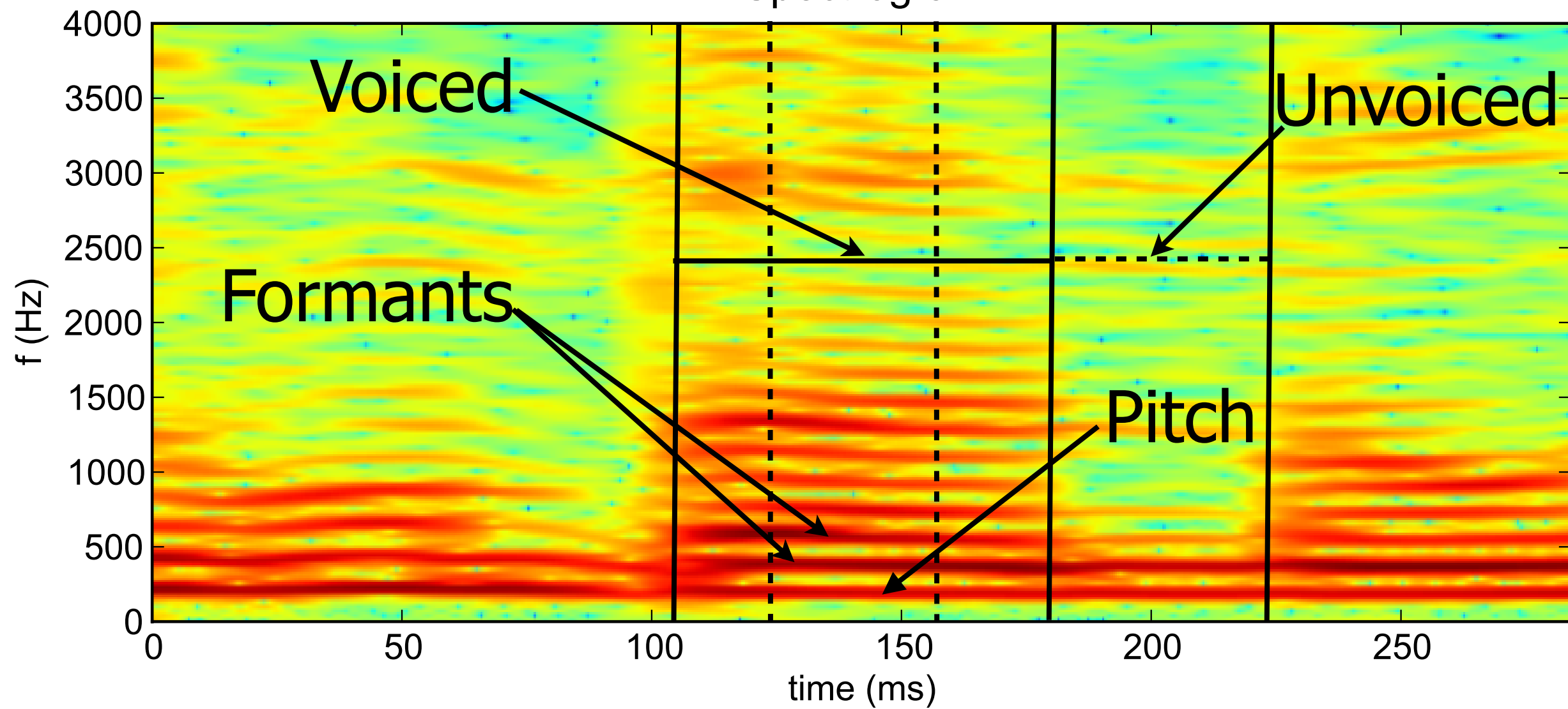
- The Big-Three of Prosody are Energy (loudness), Pitch (sound) and Tempo (speaking rate);
- Energy and Magnitude account for Energy;
- Zero Crossing Rate and Autocorrelation account for Pitch;
- The number of syllables per minute accounts for Tempo.



Computational Paralinguistics

- In general, the analysis windows are 30 ms long and the step is 10 ms, resulting into 100 feature vectors per second;
- For a given a speech sample, it is possible to estimate statistics of all features;
- The most common statistics are mean, variance, minimum, maximum, skewness, position of minimum and maximum, etc.

Spectrogram



Six basic features
(pitch, formant 1,
formant 2, energy,
voiced and unvoiced
length)

For each of the six basic
features, there are 4
statistics (mean,
minimum, maximum,
entropy)

$$\vec{x} = (x_1, x_2, \dots, x_{24})$$

The result is a 24-
dimensional feature
vectors

Outline

- Introduction
- Computational Paralinguistics
- **Trait Prediction**
- Conclusion

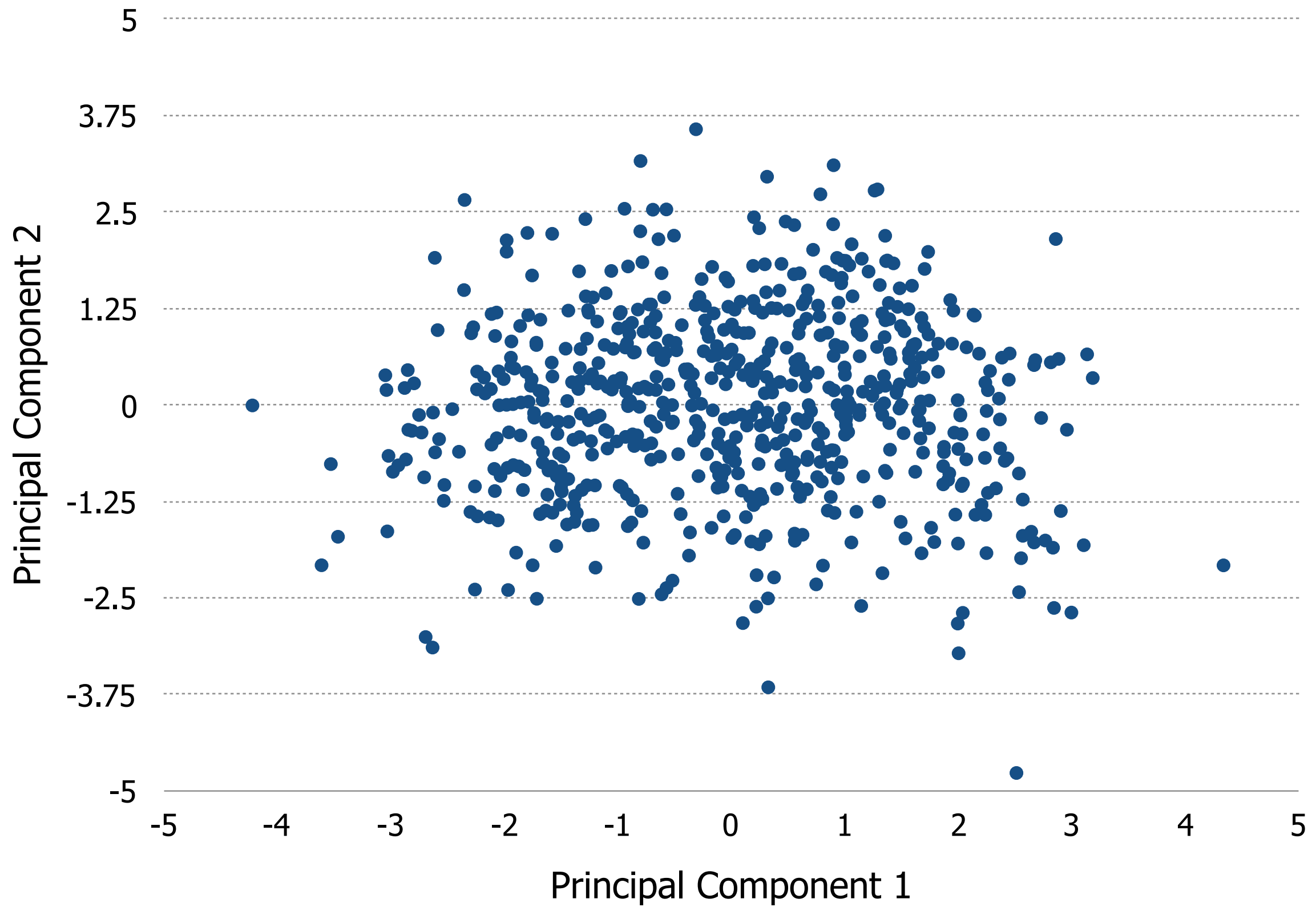


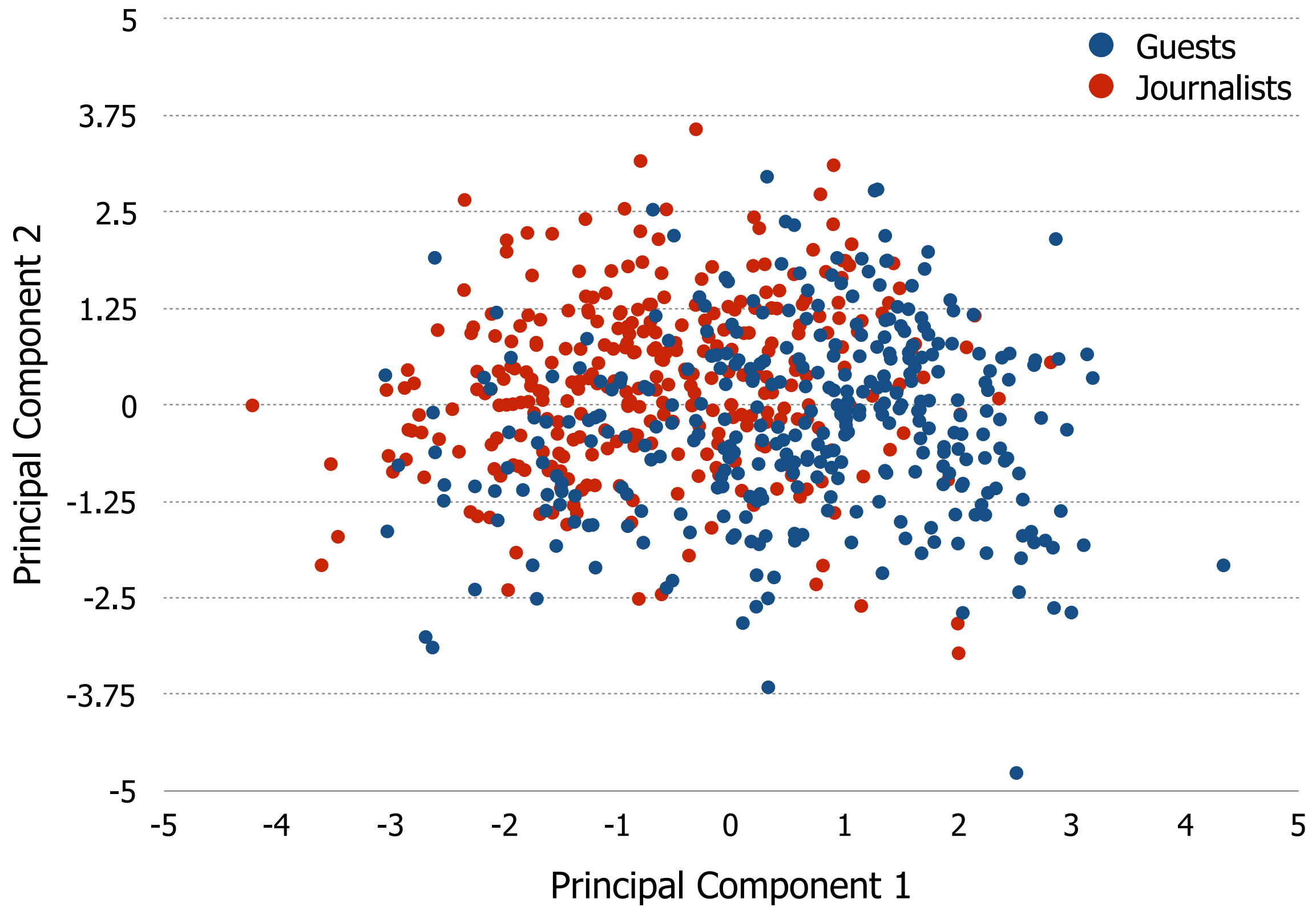
All individuals speaking in the news of Radio
Suisse Romande during February 2005

SSPNet Personality Corpus

Number of Samples	640
Total Length	1h:46m
Number of Subjects	322
Gender Balance	78.5% M / 21.5% F
Category Balance	48% J / 52% G
Speaker Distribution	80% < 3
Assessors	11 (British)
Total Items	70400

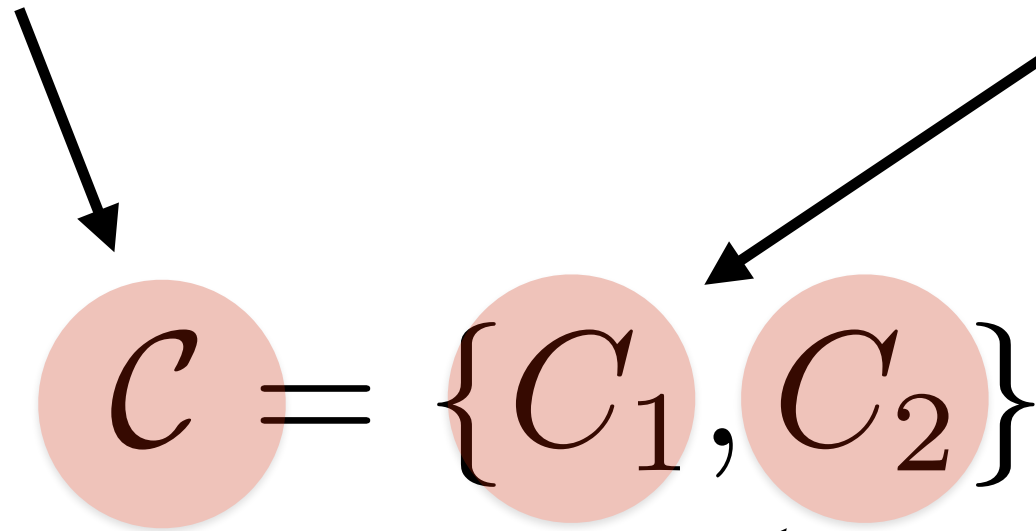
Mohammadi et al., "The Voice of Personality: Mapping Nonverbal Vocal Behavior into Trait Attributions", Social Signal Processing Workshop, pp. 17-20, 2010.





For each of the five traits, it is possible to identify two classes

Samples above median for a given trait



Samples below median for a given trait

Class assigned
automatically to a
speech sample

Vector extracted from
the speech sample


$$C^* = \arg \max_{C \in \mathcal{C}} p(C | \vec{x}) p(C)$$

Posterior of the class

Prior of the class
(uninformative in this
case)

The result is a 24-dimensional feature vectors

Scalar product between a parameter vector and the feature vector

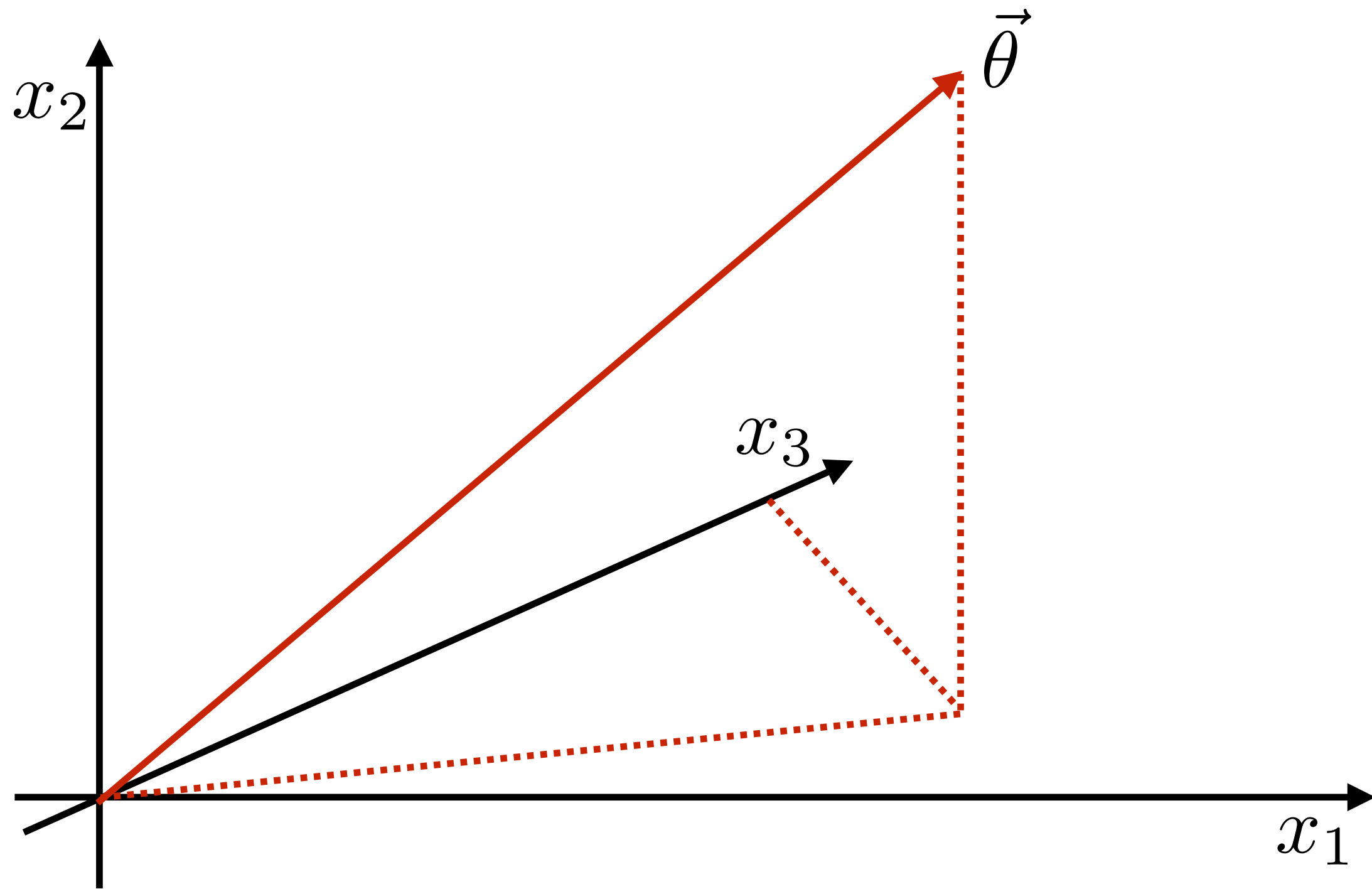
The diagram illustrates the logistic regression equation with several annotations. An arrow points from the text 'The result is a 24-dimensional feature vectors' to the feature vector \vec{x} in the probability term $p(C|\vec{x})$. Another arrow points from the text 'Scalar product between a parameter vector and the feature vector' to the term $\vec{\theta} \vec{x}$ in the exponent. A third arrow points from the text 'A scalar threshold' to the term θ_0 in the exponent. The equation is as follows:

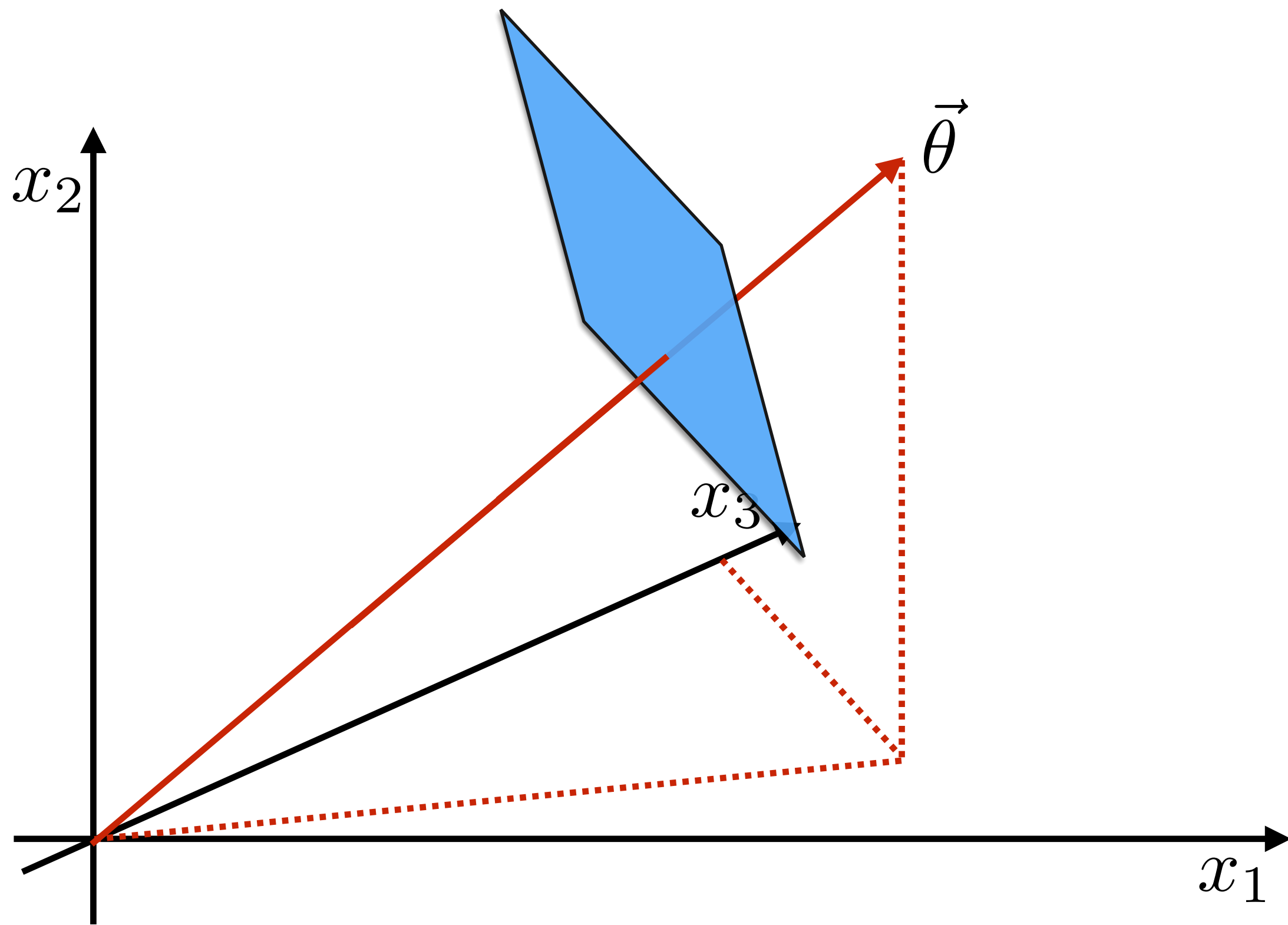
$$p(C|\vec{x}) = \frac{\exp(\vec{\theta} \vec{x} - \theta_0)}{1 + \exp(\vec{\theta} \vec{x} - \theta_0)}$$

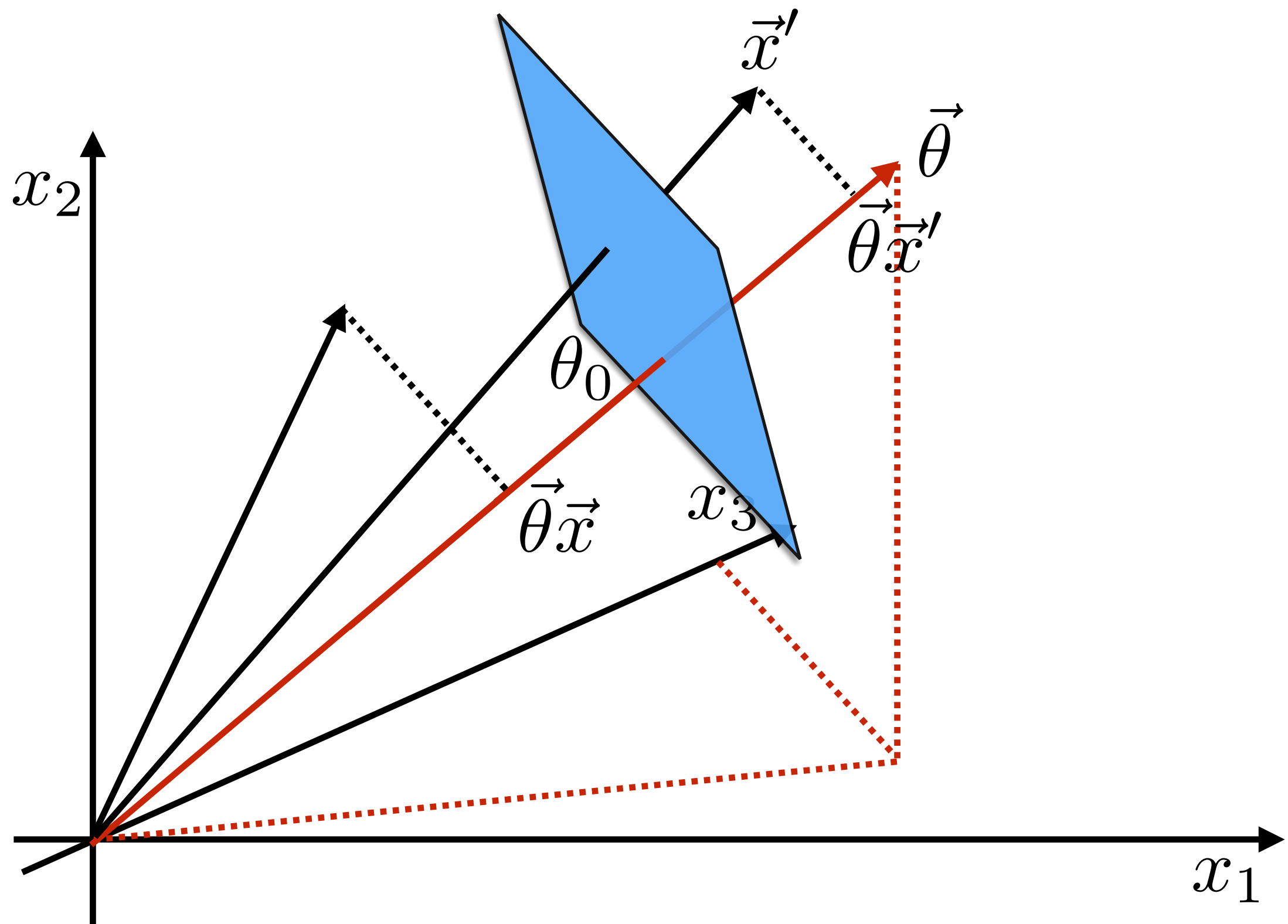
A scalar threshold

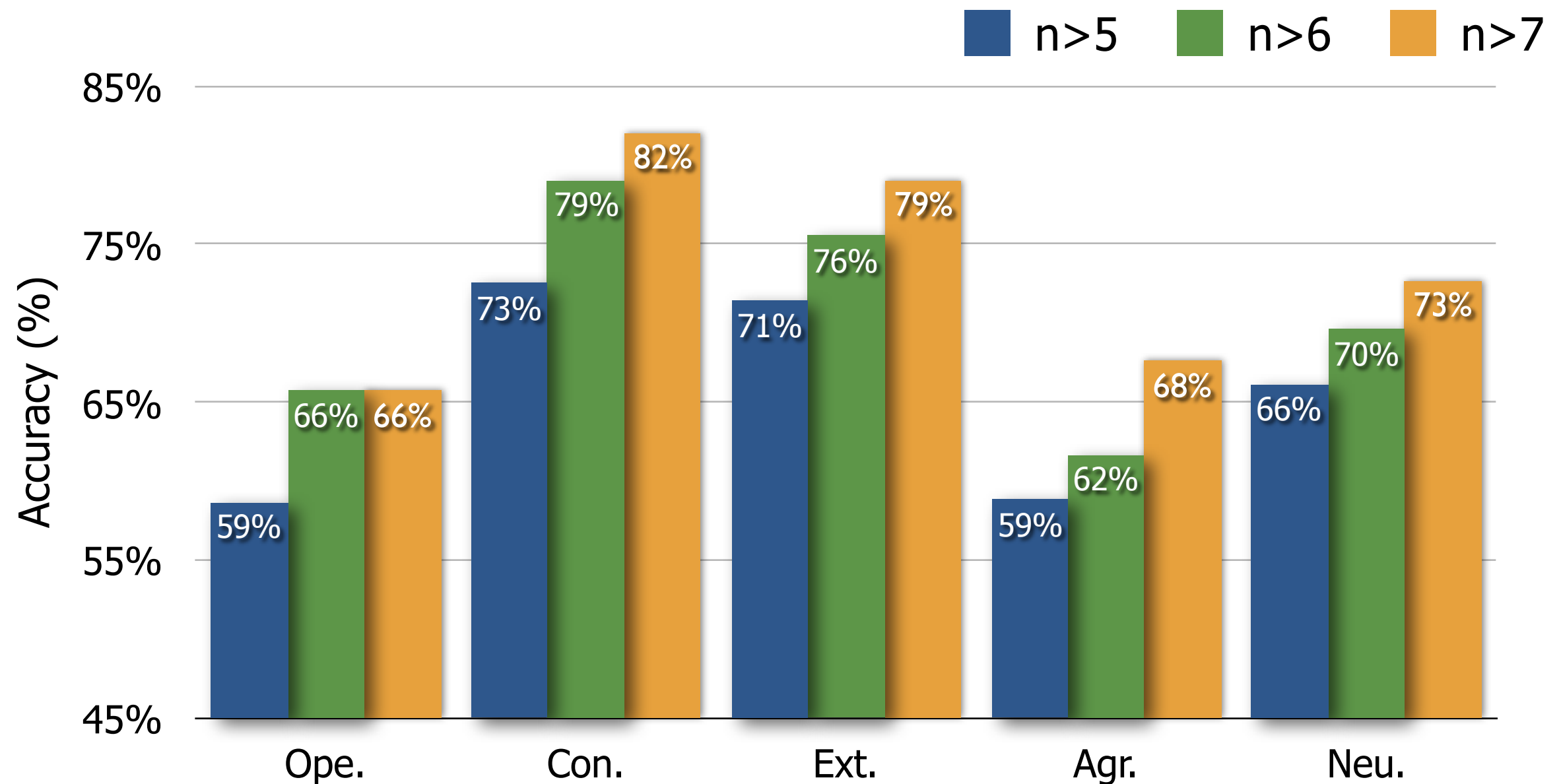
Training

- In the case of the logistic regression, the training aims at finding the parameters (the components of theta and the threshold);
- The training is performed by iteratively changing the values of the parameters to minimise the error rate;
- Such a process is performed using a k-fold approach.









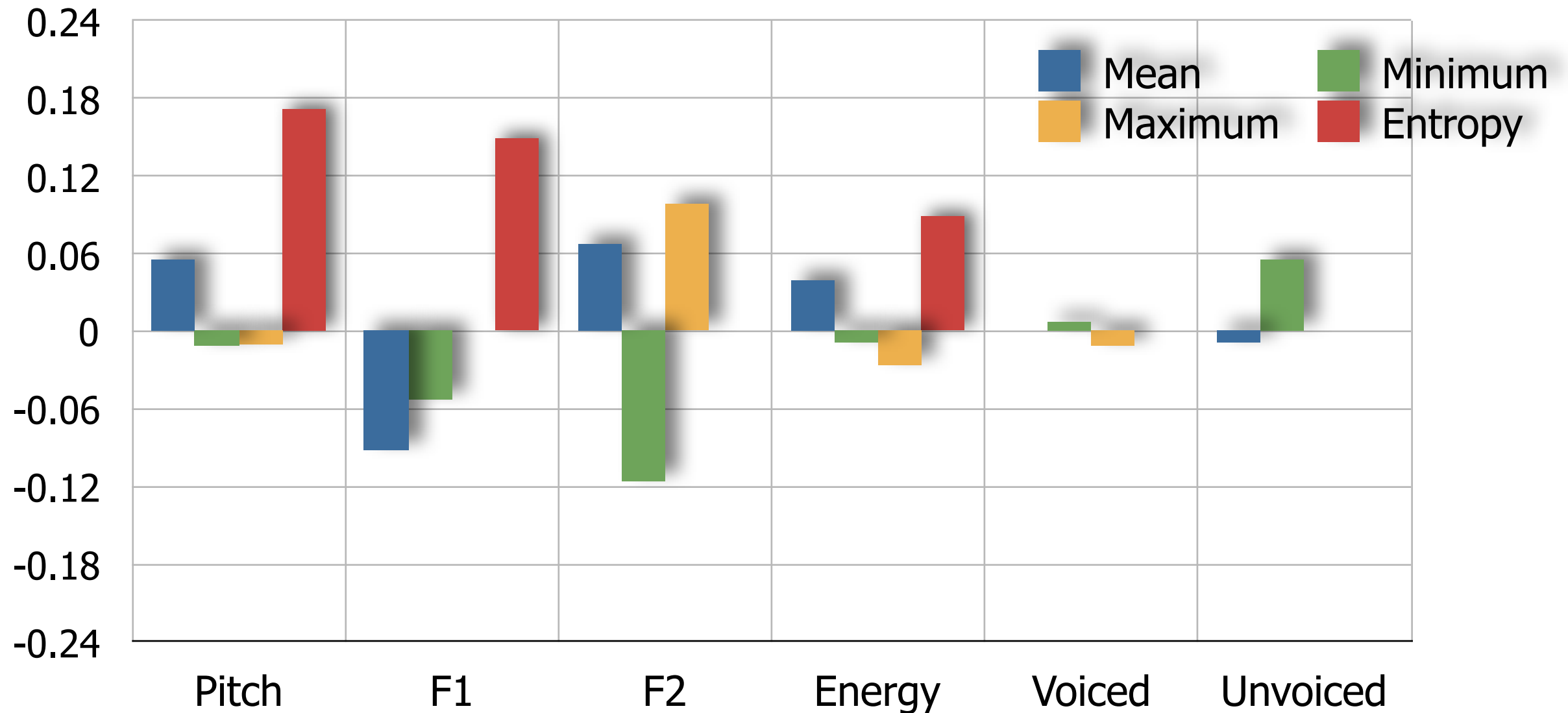
Mohammadi & Vinciarelli, "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features", IEEE Transactions on Affective Computing, 3(3):273-284, 2012

Speech and Personality

“[...] there are two dimensions that underlie most judgments of traits, people, groups, and cultures [...] the first makes reference to attributes such as competence [...] and the second to warmth [...]”

Judd et al., “Fundamental Dimensions of Social Judgment: Understanding the Relations Between Judgments of Competence and Warmth”, Journal of Personality and Social Psychology, 89(6):899-913, 2005

Conscientiousness



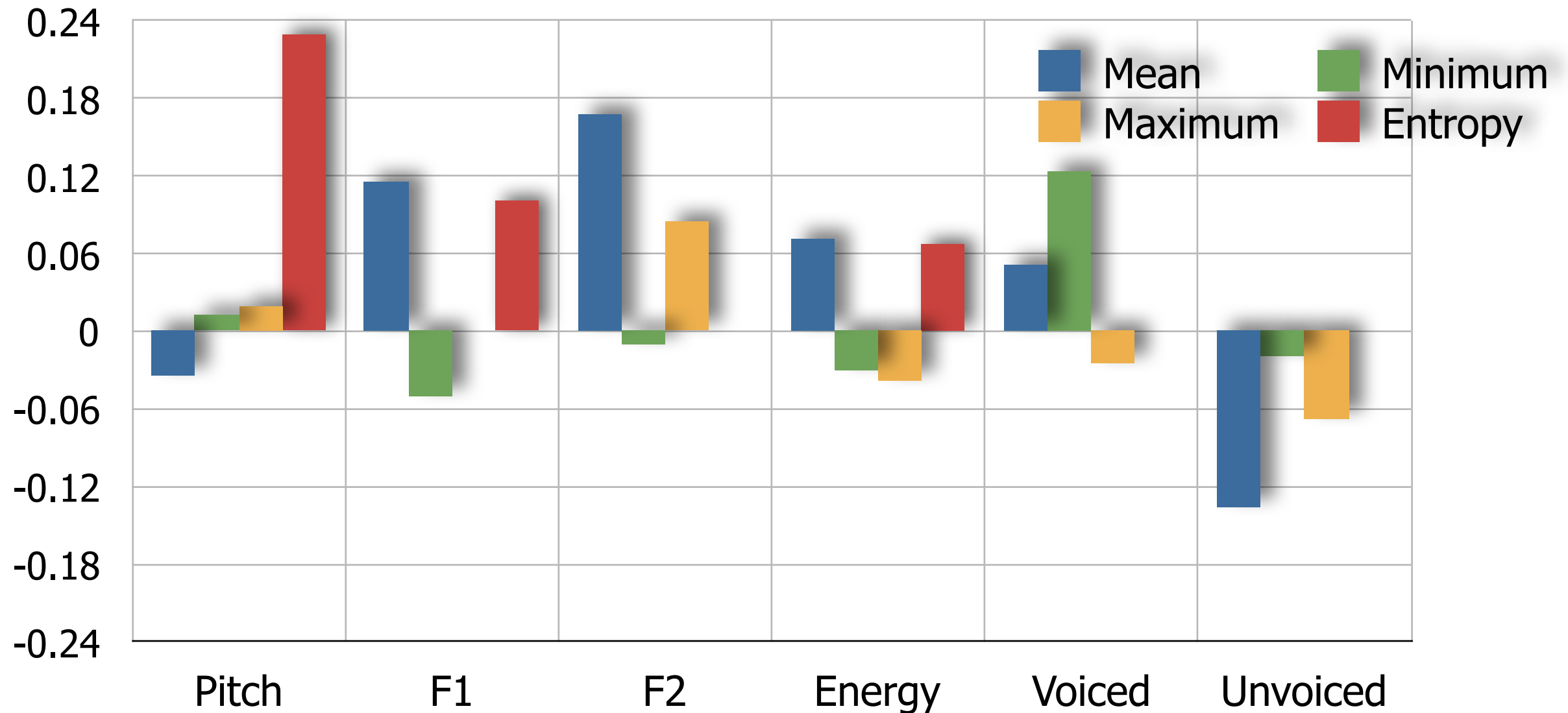
Mohammadi & Vinciarelli, "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features", IEEE Transactions on Affective Computing, 3(3):273-284, 2012

Speech and Personality

“Rate and pitch variation were the most influential for competence and benevolence, respectively. For competence, one interaction effect (rate by pitch variation) was significant.”

Ray, “Vocally cued personality prototypes: An implicit personality theory approach”, Journal of Communication Monographs, 53(3):266-276, 1986

Extraversion



Mohammadi & Vinciarelli, "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features", IEEE Transactions on Affective Computing, 3(3):273-284, 2012

Speech and Personality

“Rate and pitch variation were the most influential for competence and benevolence, respectively. For competence, one interaction effect (rate by pitch variation) was significant.”

Ray, “Vocally cued personality prototypes: An implicit personality theory approach”, Journal of Communication Monographs, 53(3):266-276, 1986

Outline

- Introduction
- Computational Paralinguistics
- Trait Prediction
- Conclusion

Conclusions

- Speech signals can be analysed in the time domain through convolution operations;
- In most cases, the processing takes place in the frequency domain (after performing Fourier transform);
- The main reason for analysing speech is that it is the main form of communication between people.