Time Dependent ARMA for Automatic Recognition of Fear-Type Emotions in Speech

J.C. Vásquez-Correa[1], J.R Orozco-Arroyave[1,2], J.D Arias-Londoño[1], J.F Vargas-Bonilla[1], L.D. Avendaño[3], Elmar Nöth[2]

[1] Faculty of Engineering, University of Antioquia UdeA
[2] Pattern Recognition Lab, Friedrich Alexander Universität, Erlangen-Nürnberg
[3] Mechanical & Aeronautical Engineering, University of Patras

Text, Speech and Dialogue, TSD 2015

Outline

1. **Introduction**
2. Methodology
3. Databases and experiments
4. Results
5. Conclusion

1. Introduction

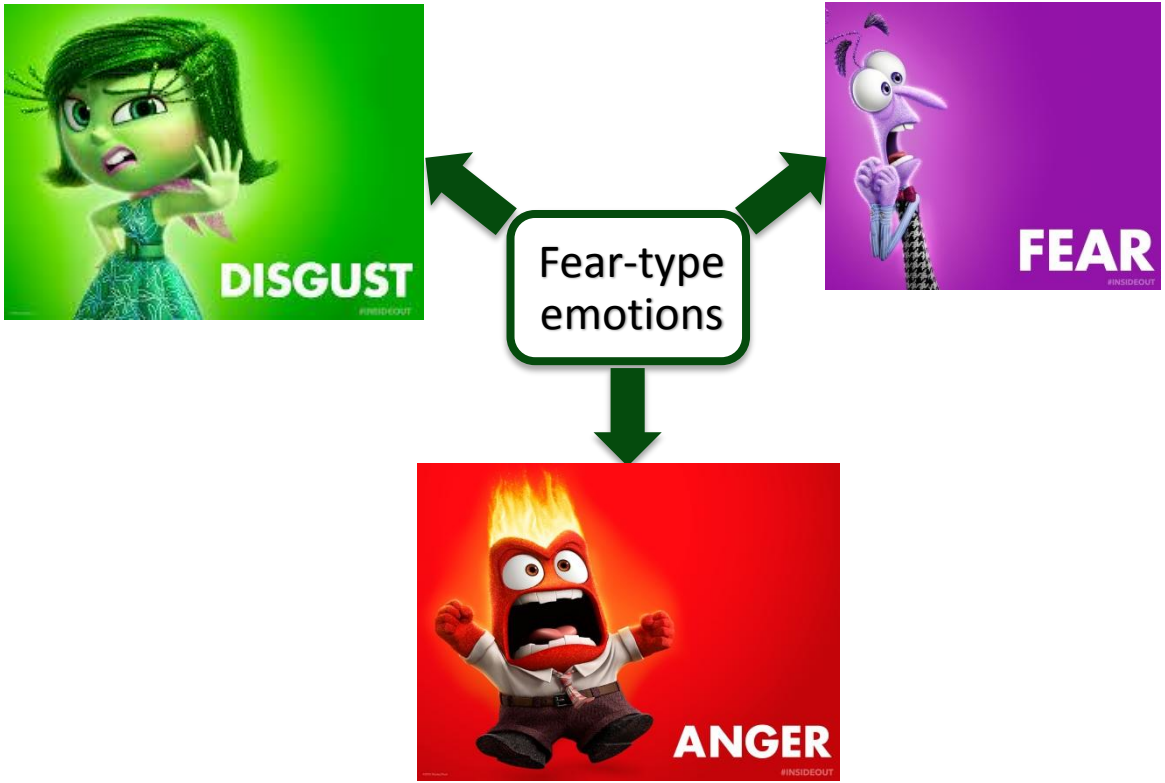
Recognition of emotion in speech:

- ✓ Call centers
- ✓ Emergency services
- ✓ psychological therapy
- ✓ Intelligent vehicles
- ✓ Surveillance

UNIVERSIDAD DE ANTIOQUIA
1803

jcamilo.vasquez@udea.edu.co

1. Introduction

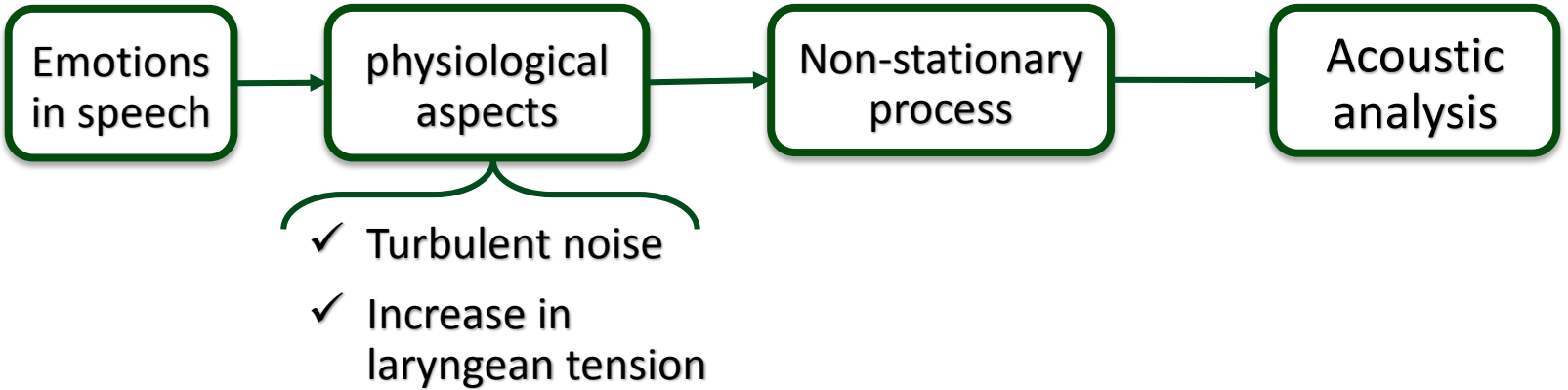


UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

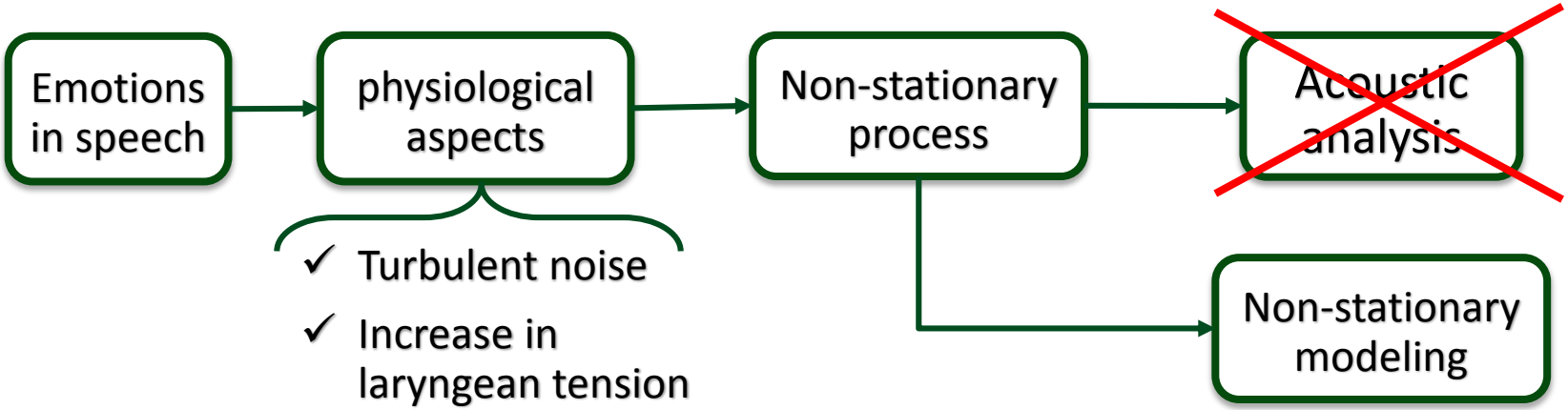
1. Introduction



UNIVERSIDAD DE ANTIOQUIA
1803

FAU

1. Introduction

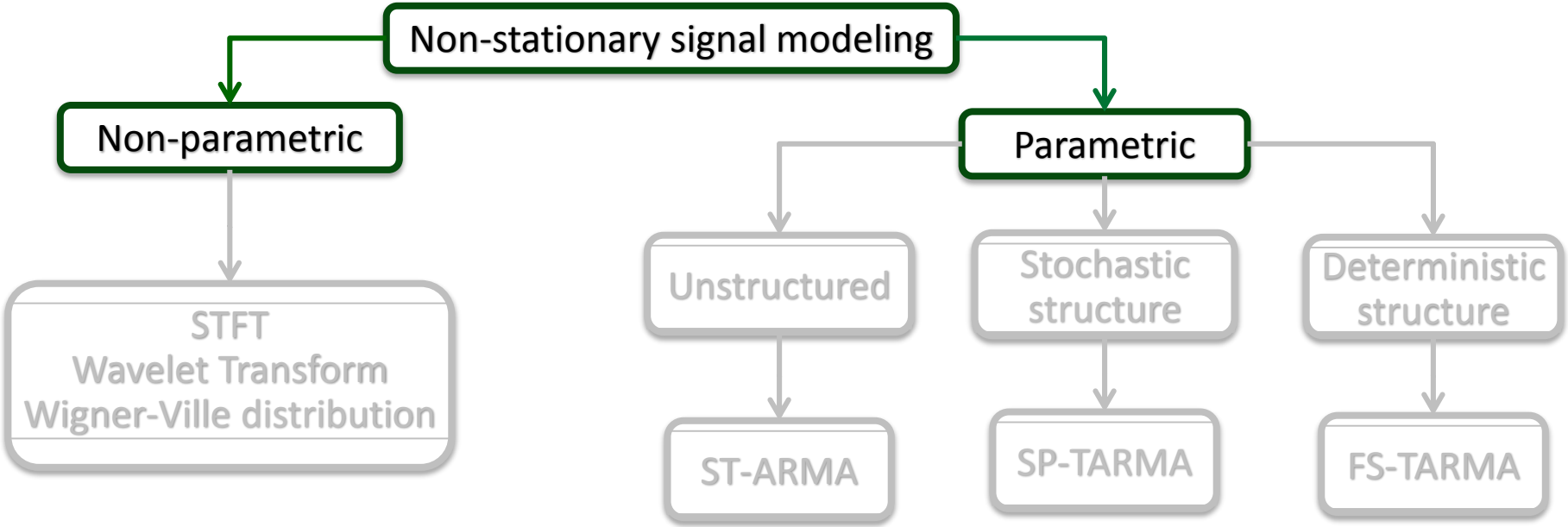


UNIVERSIDAD DE ANTIOQUIA
1803

FAU

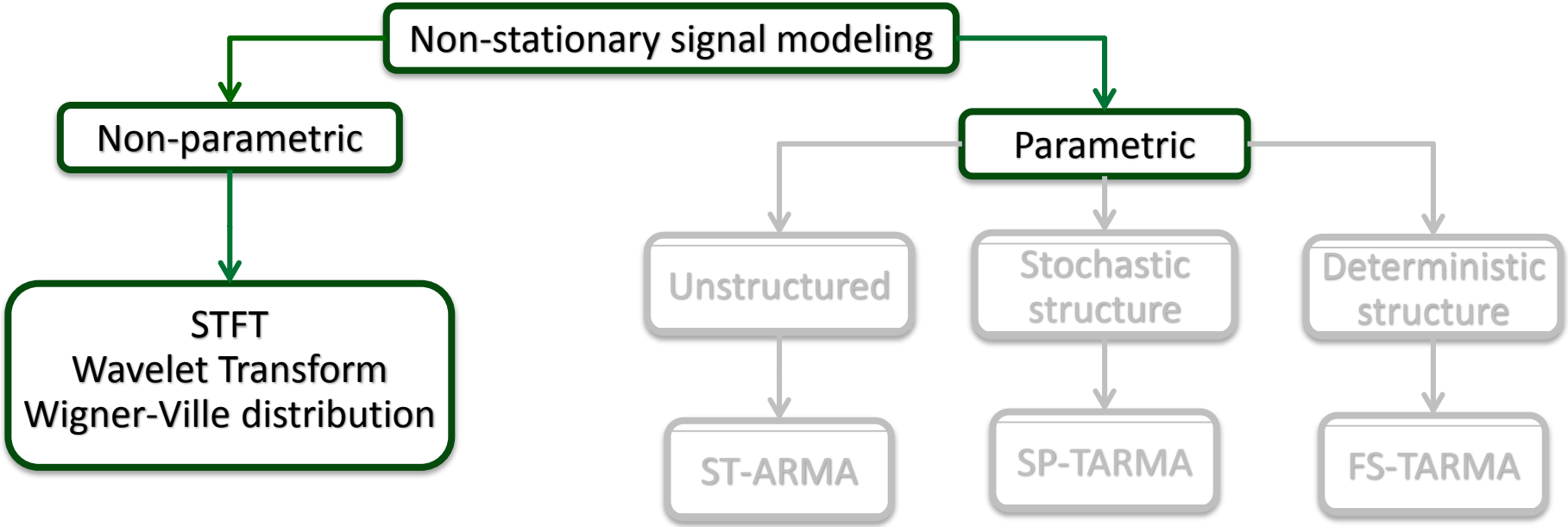
jcamilo.vasquez@udea.edu.co

1. Introduction



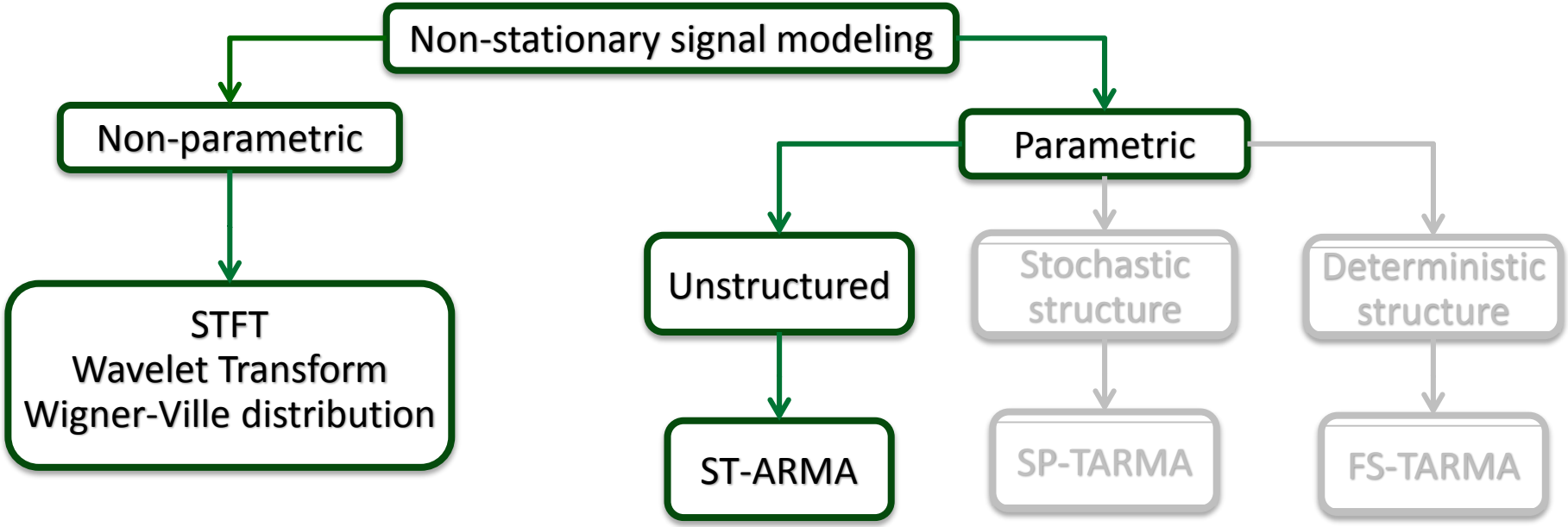
jcamilo.vasquez@udea.edu.co

1. Introduction



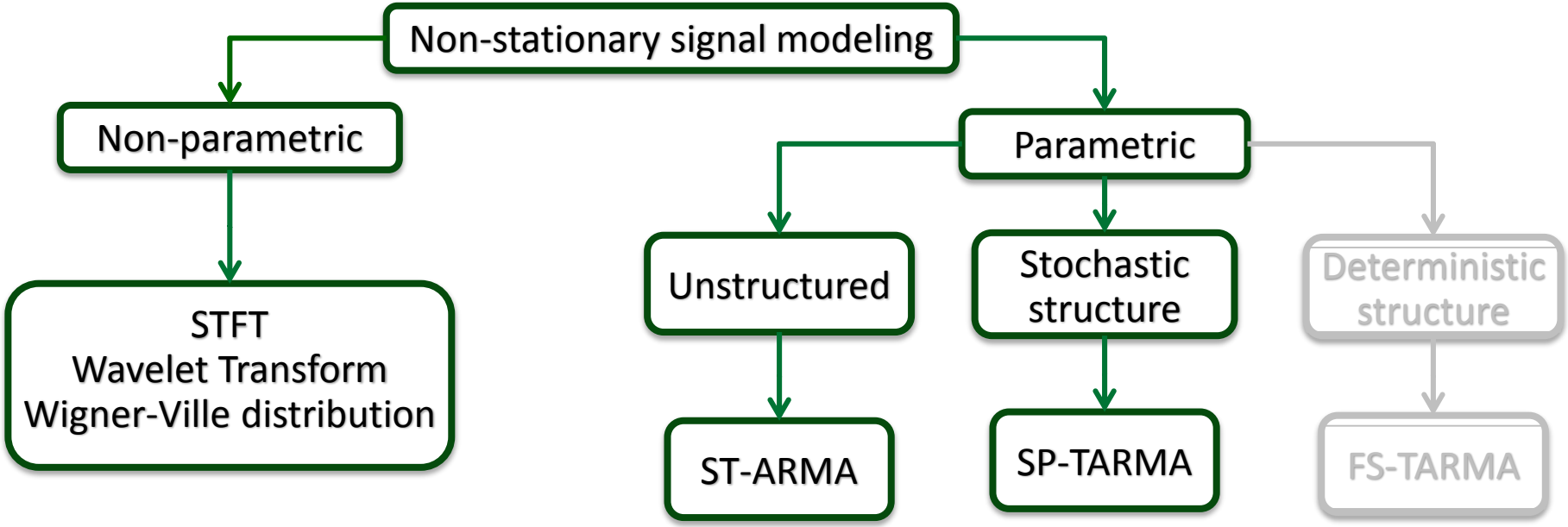
jcamilo.vasquez@udea.edu.co

1. Introduction

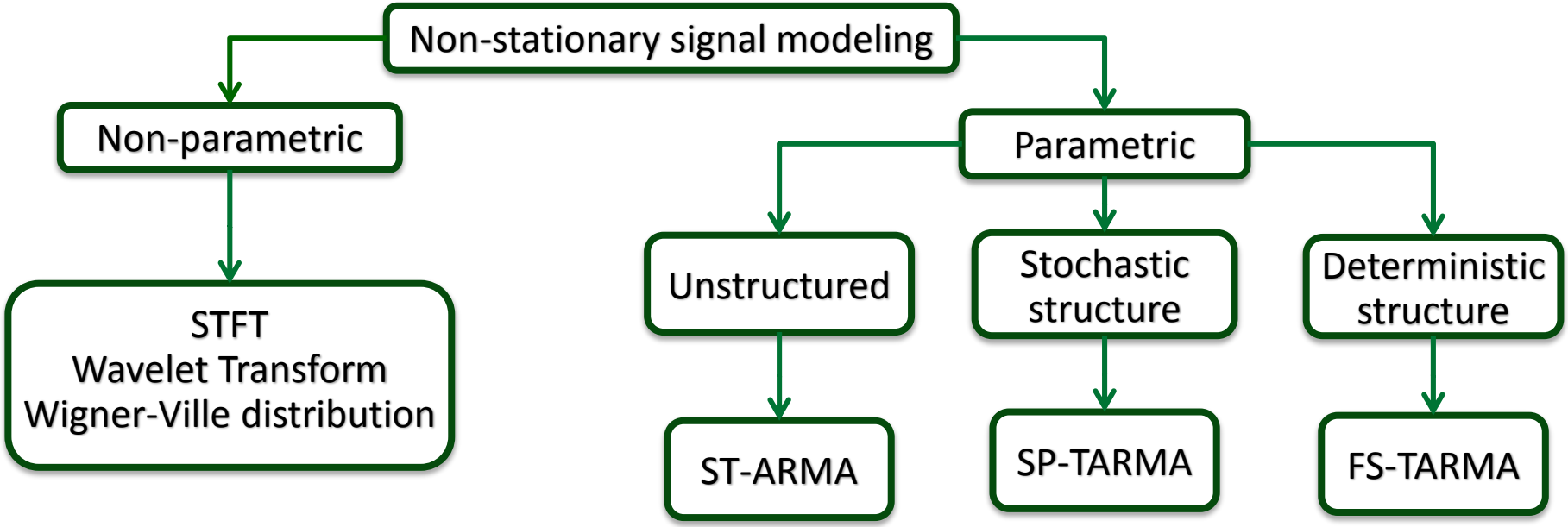


jcamilo.vasquez@udea.edu.co

1. Introduction



1. Introduction

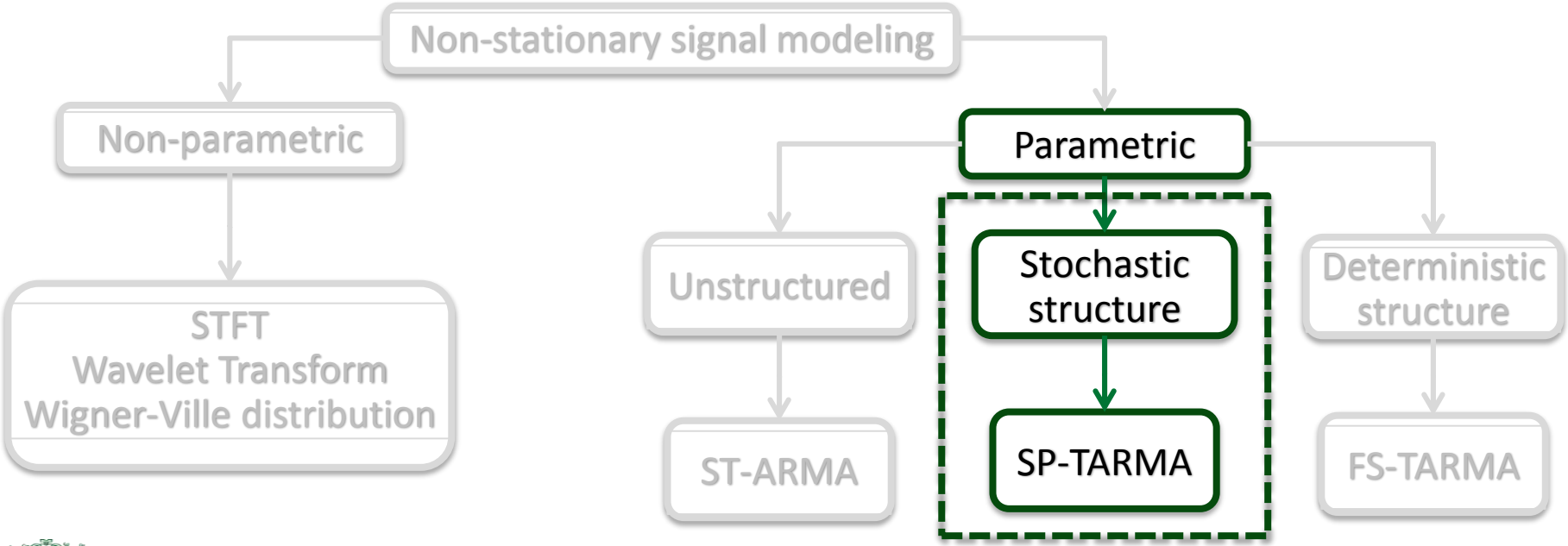


UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

1. Introduction

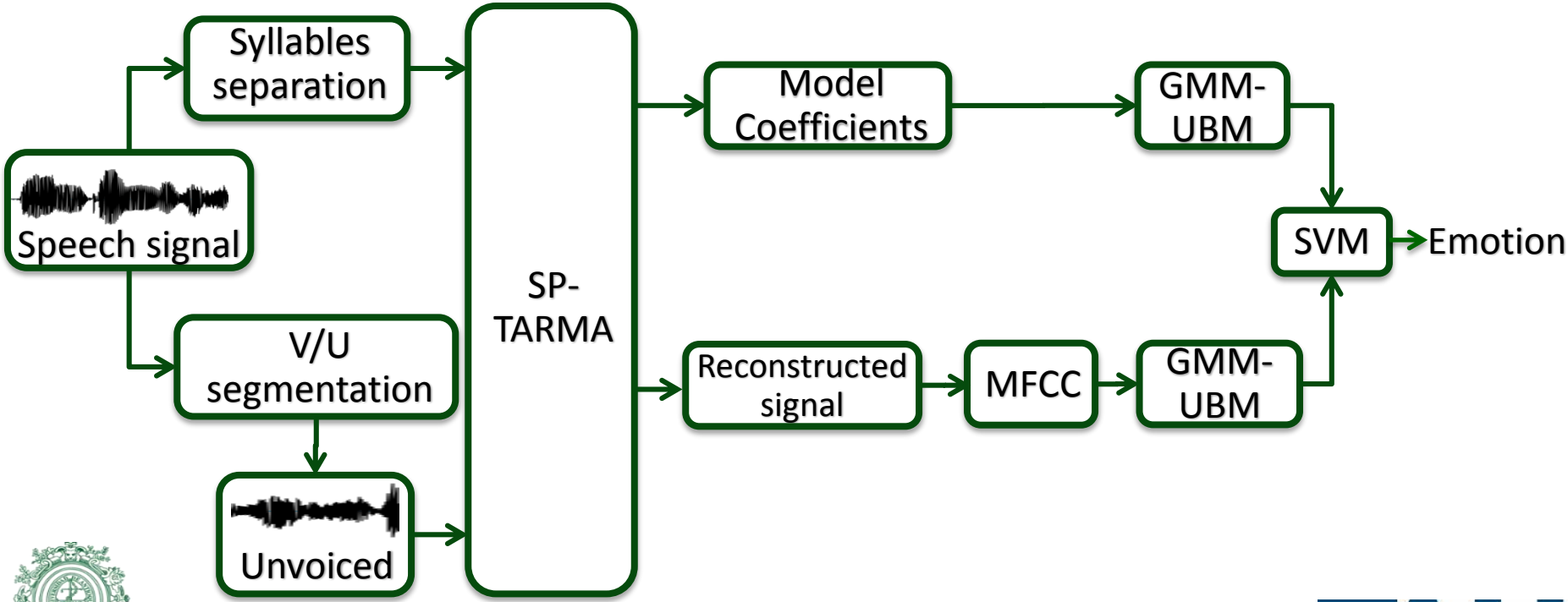


jcamilo.vasquez@udea.edu.co

Outline

1. Introduction
2. **Methodology**
3. Databases and experiments
4. Results
5. Conclusion

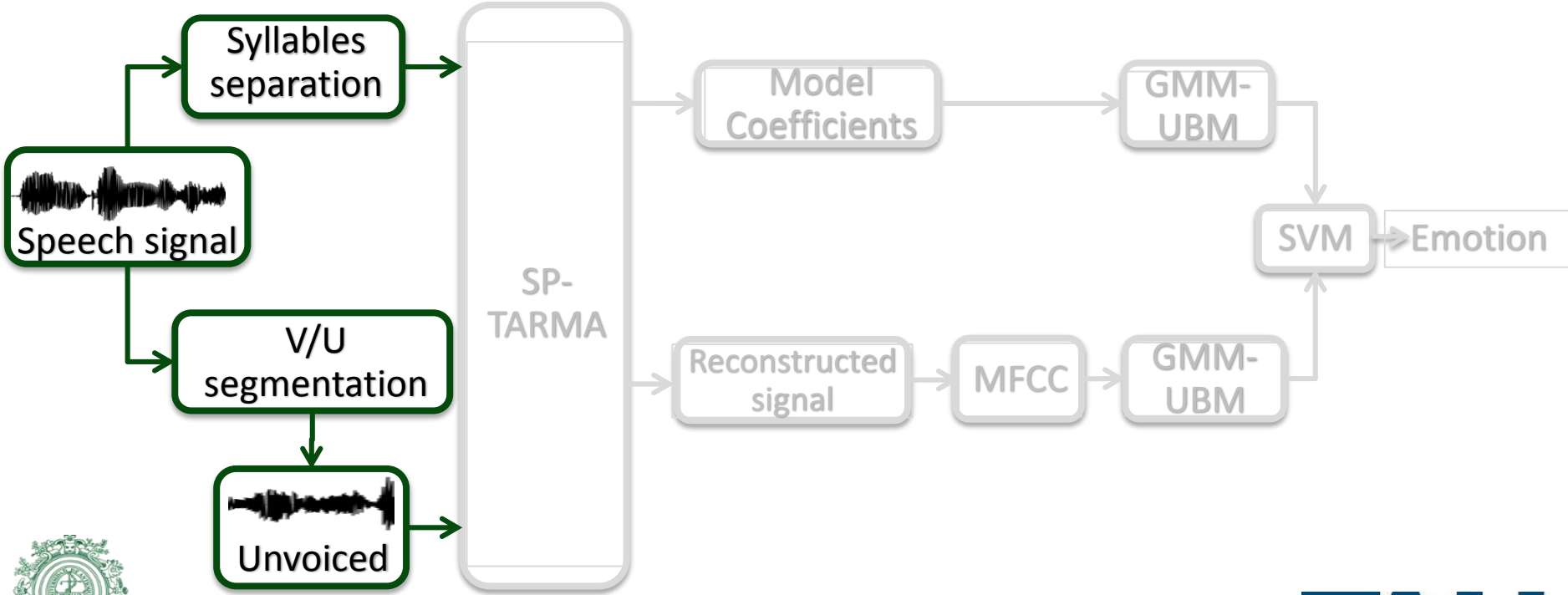
2. Methodology



UNIVERSIDAD DE ANTIOQUIA
1803

FAU

2. Methodology

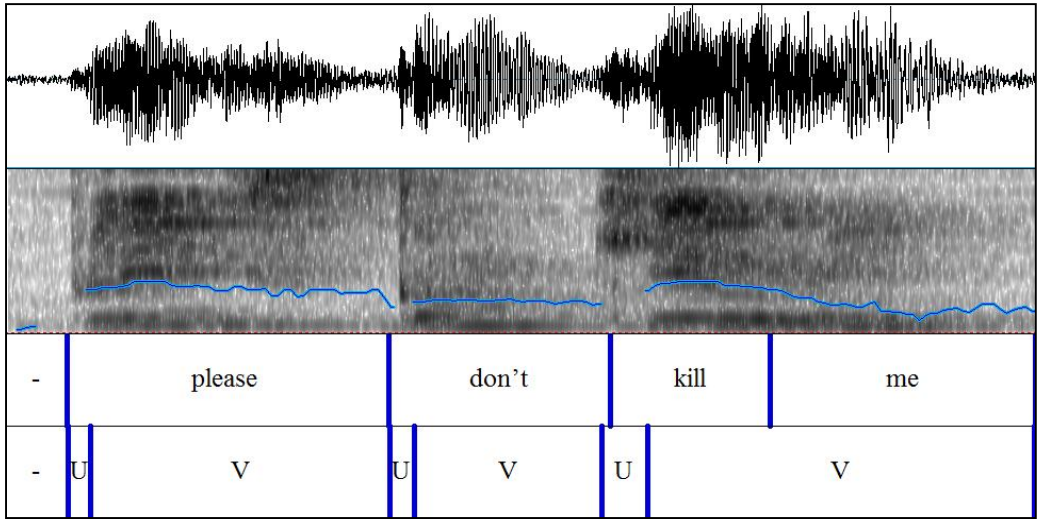


2. Methodology

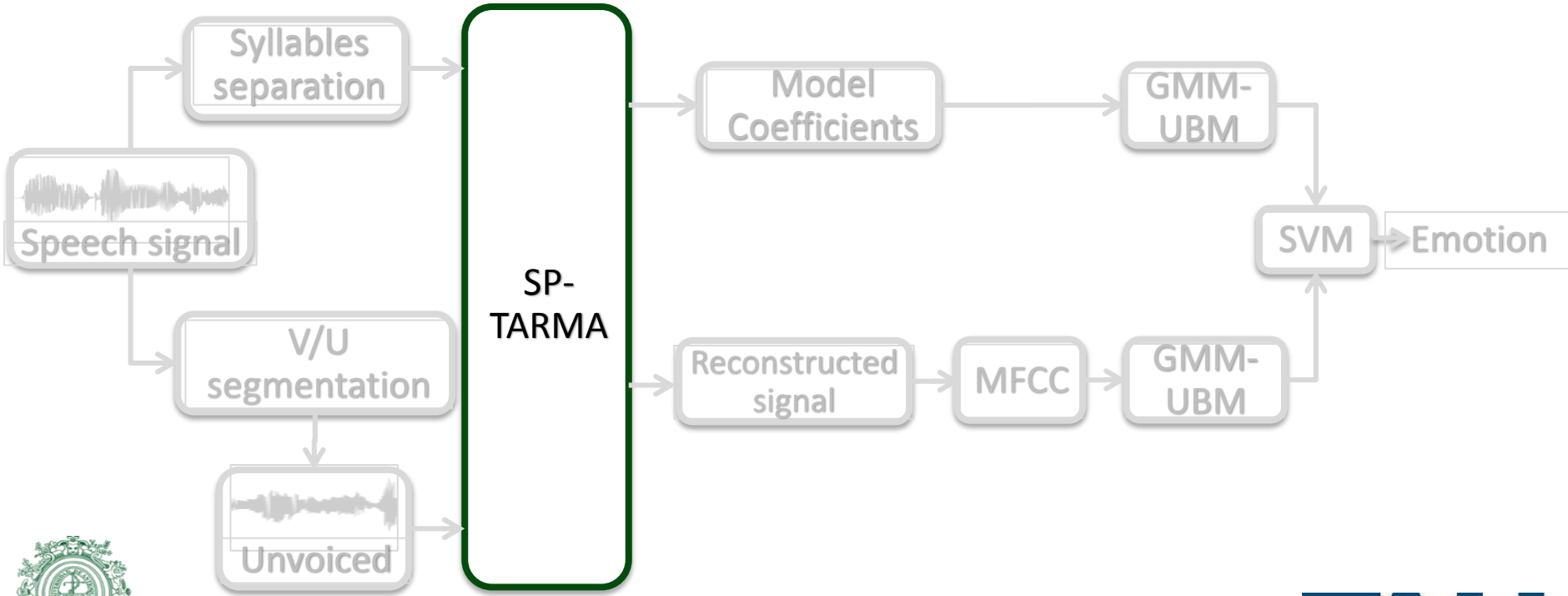
Segmentation

Two types of segments are analyzed:

- ✓ Syllables
- ✓ Unvoiced

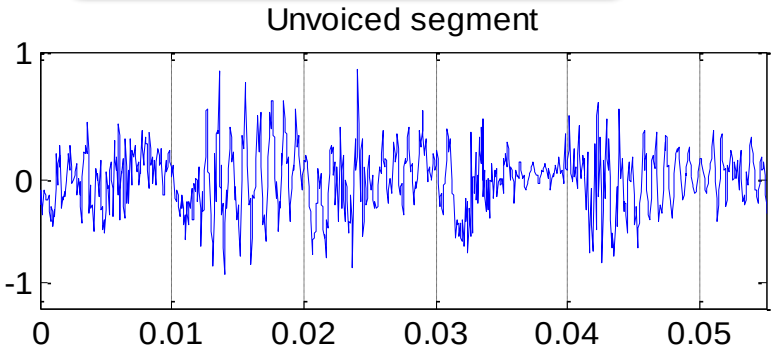


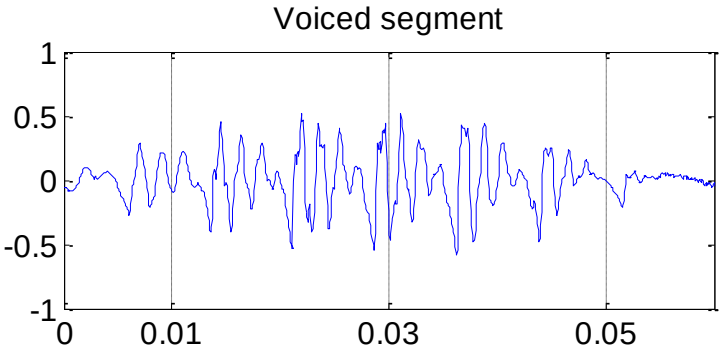
2. Methodology

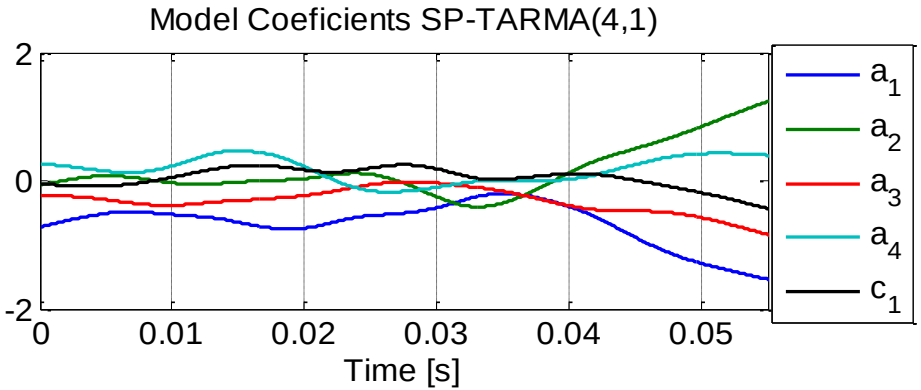


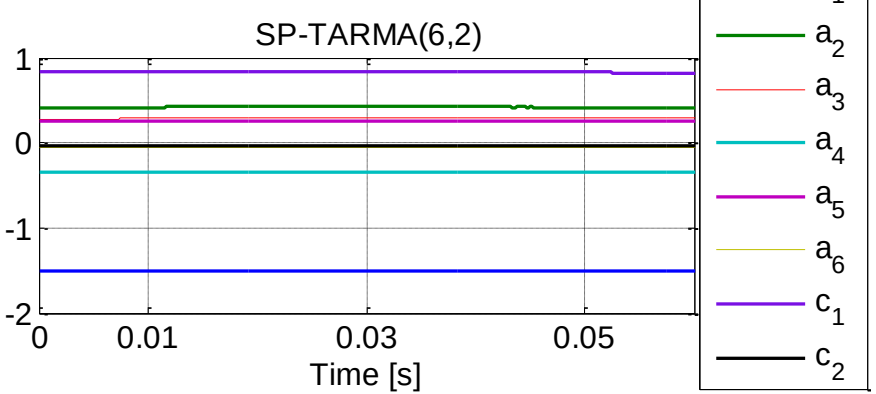
2. Methodology

SP-TARMA modeling



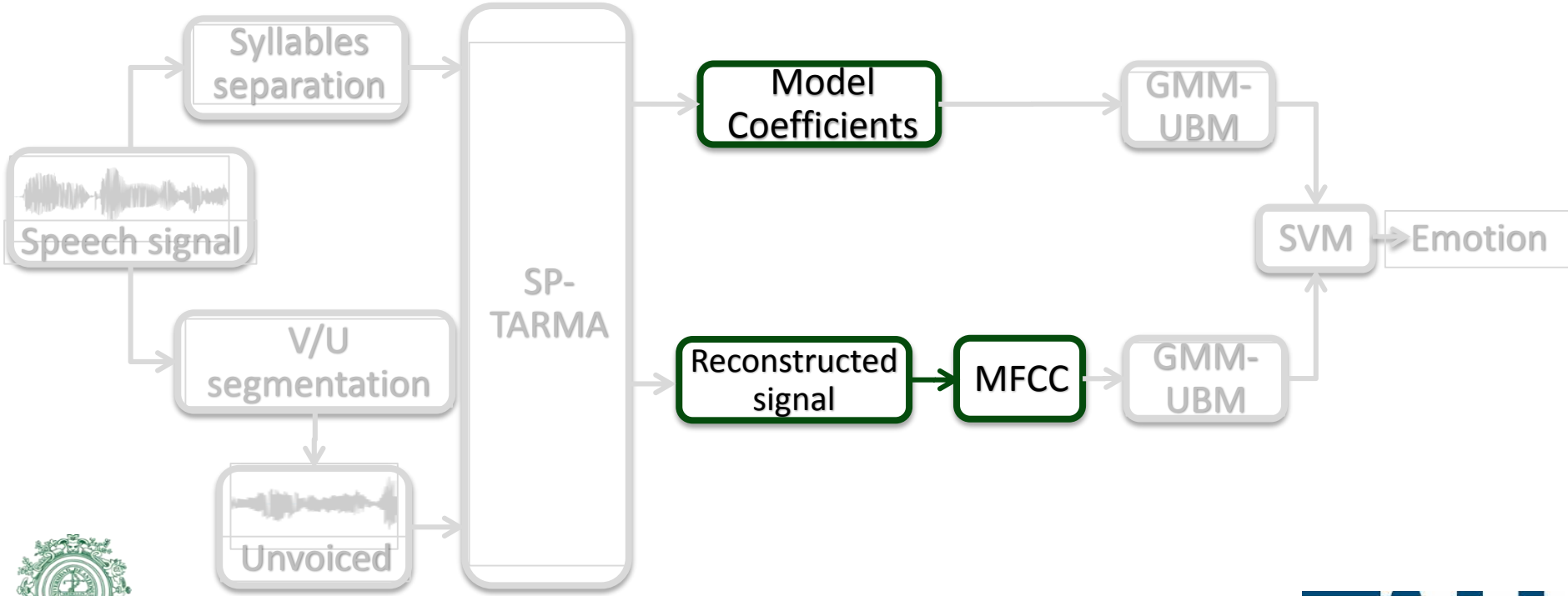






jcamilo.vasquez@udea.edu.co

2. Methodology



2. Methodology

Characterization

Feature set 1

From model Coefficients:

- ✓Mean
- ✓Stadard deviation
- ✓Skewness
- ✓Kurtosis
- ✓Maximum
- ✓Minimum
- ✓Log-energy

Feature set 2

From reconstructed signal:

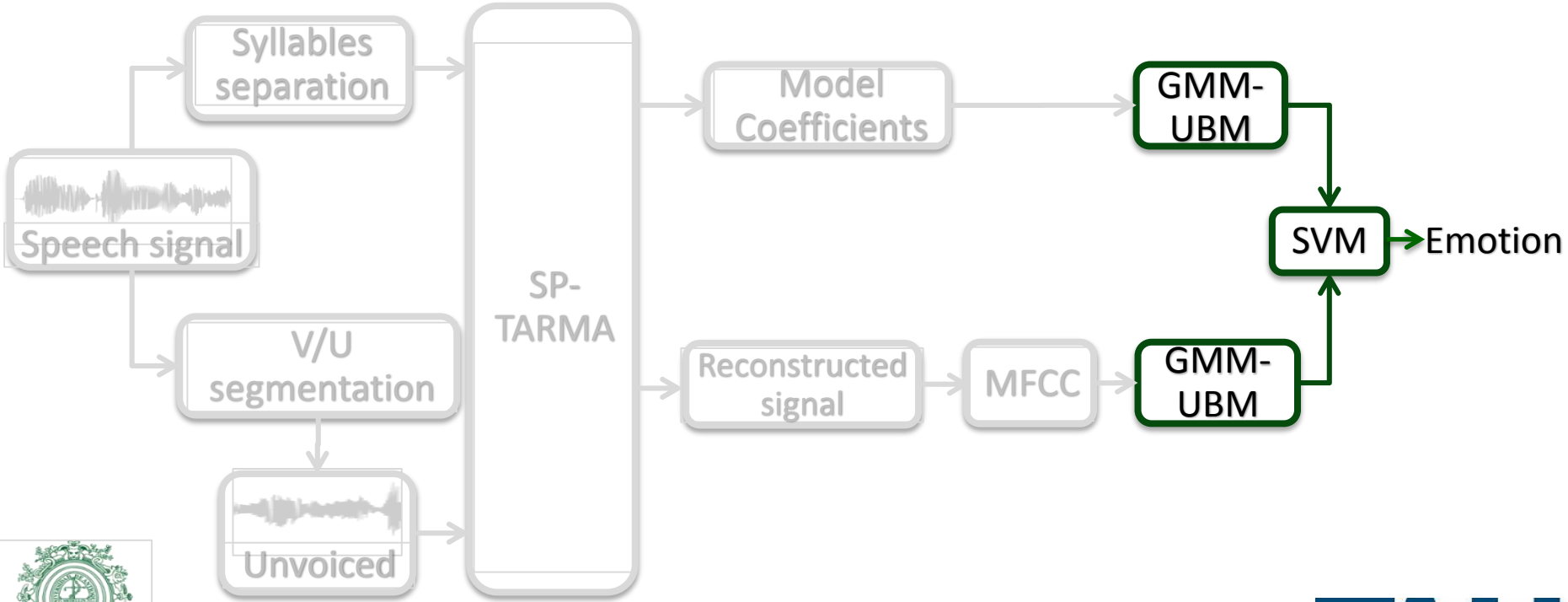
- ✓ 12 MFCC

UNIVERSIDAD DE ANTIOQUIA
1803

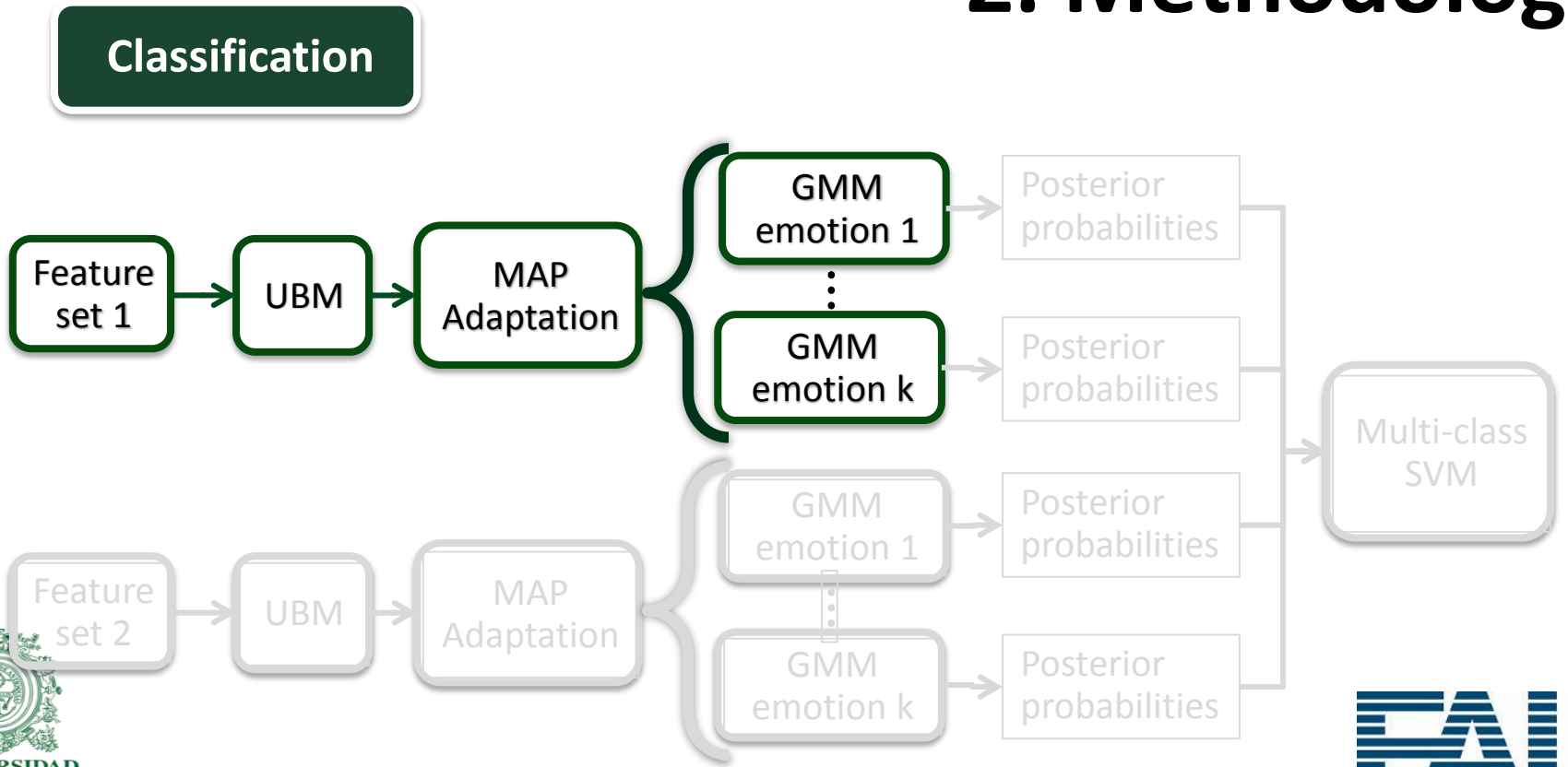
FAU

jcamilo.vasquez@udea.edu.co

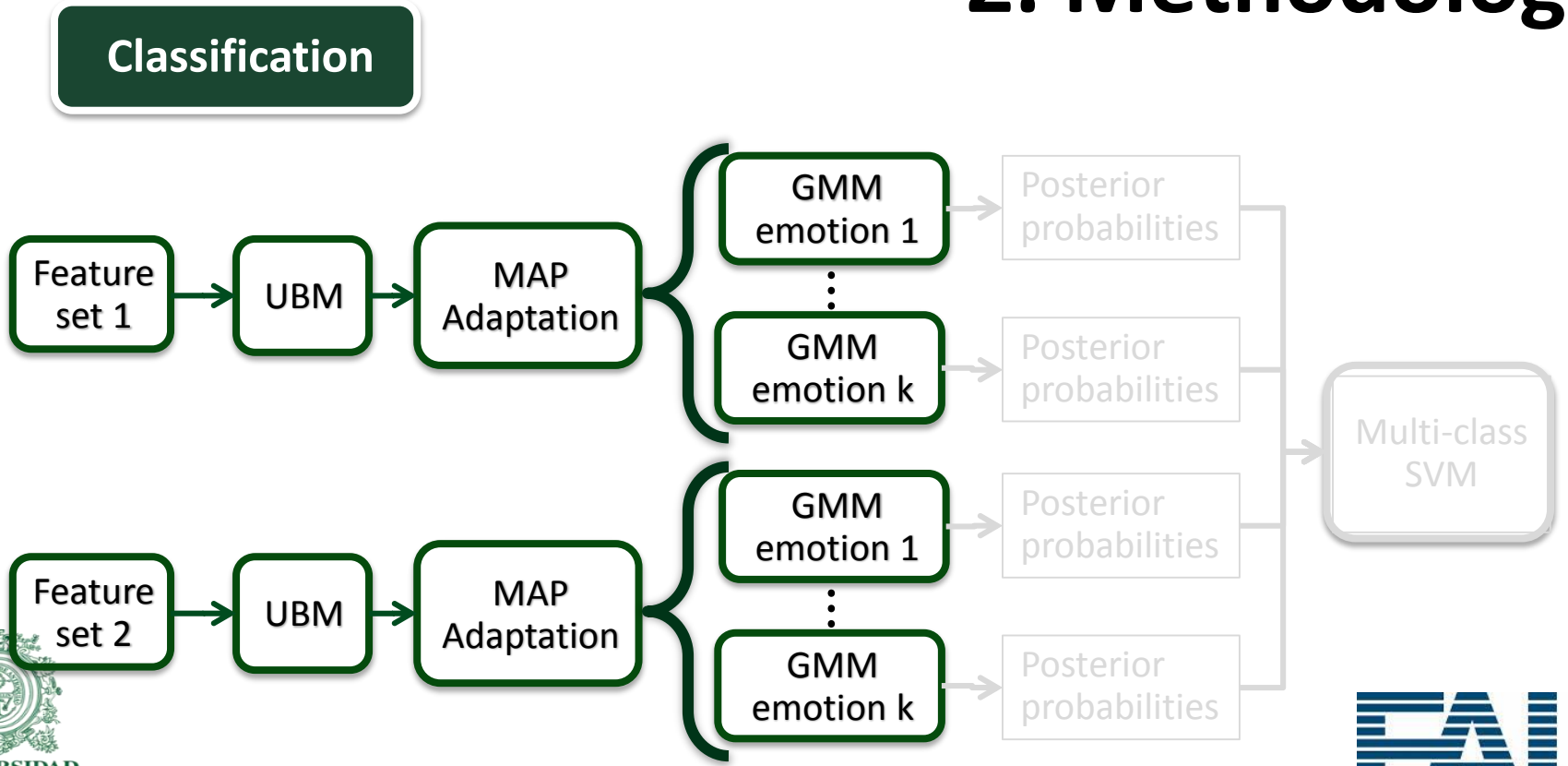
2. Methodology



2. Methodology



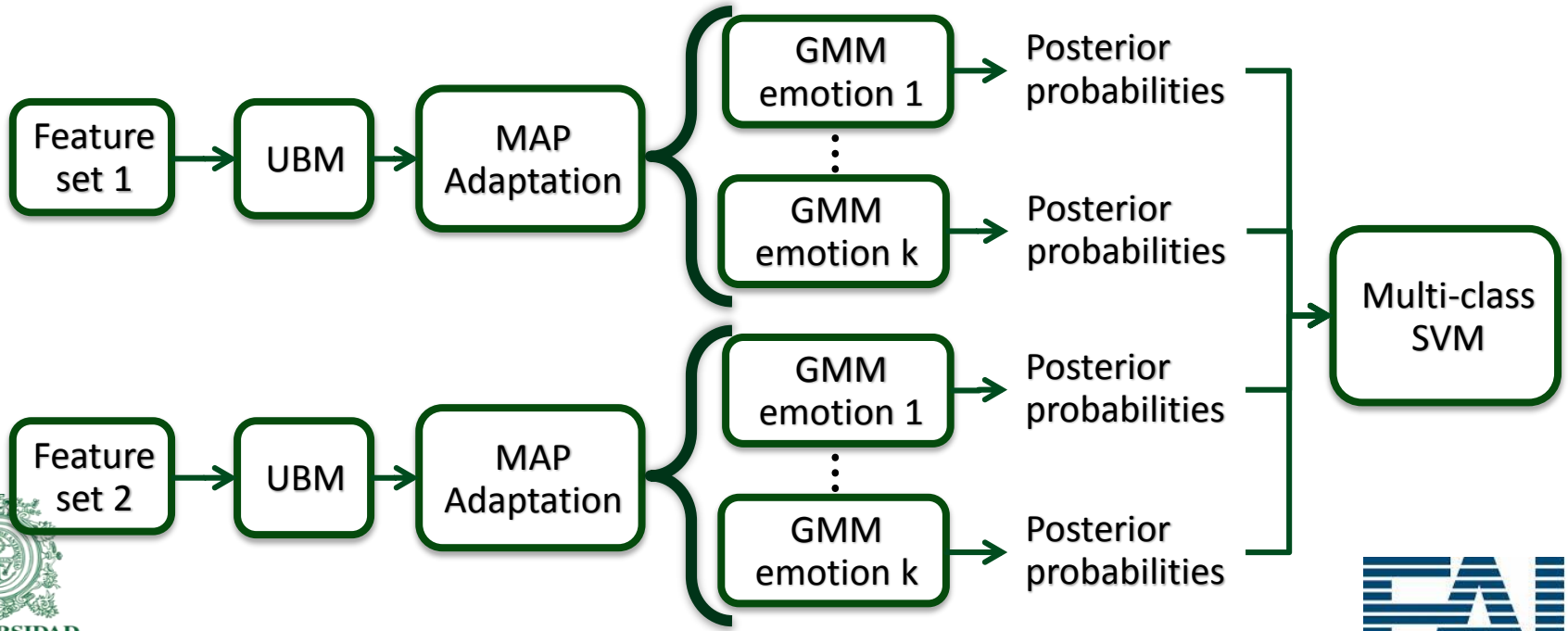
2. Methodology



jcamilo.vasquez@udea.edu.co

2. Methodology

Classification



jcamilo.vasquez@udea.edu.co

2. Methodology

Optimization parameters

- ✓ Leave One Speaker Out Cross-validation
- ✓ # of Gaussian components: 2-8
- ✓ Diagonal covariance matrix
- ✓ $10^{-1} < C < 10^{4}$
- ✓ $10^{-2} < \gamma < 10^{2}$

UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

Outline

1. Introduction
2. Methodology
3. **Databases and experiments**
4. Results
5. Conclusion

jcamilo.vasquez@udea.edu.co

3. Databases

Database	# recordings	# speakers	Sampling frequency
Berlin	534	10	16000 Hz
Enterface05	1317	44	44100 Hz

3. Experiments

Multi-class

Berlin DB

Enterface05 DB

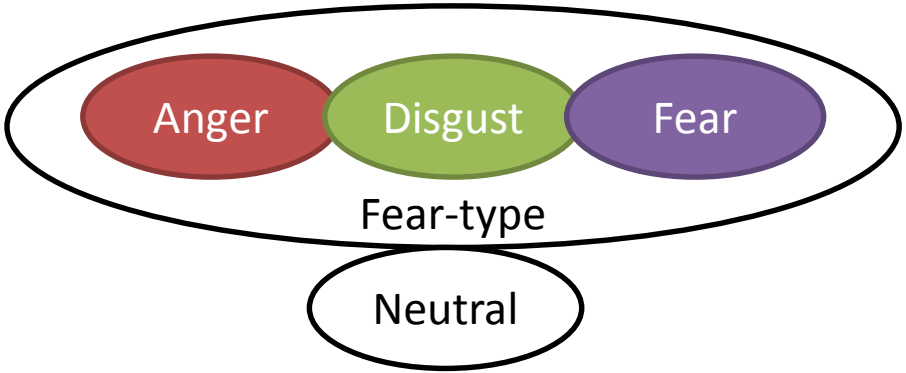
UNIVERSIDAD DE ANTIOQUIA
1803

FAU

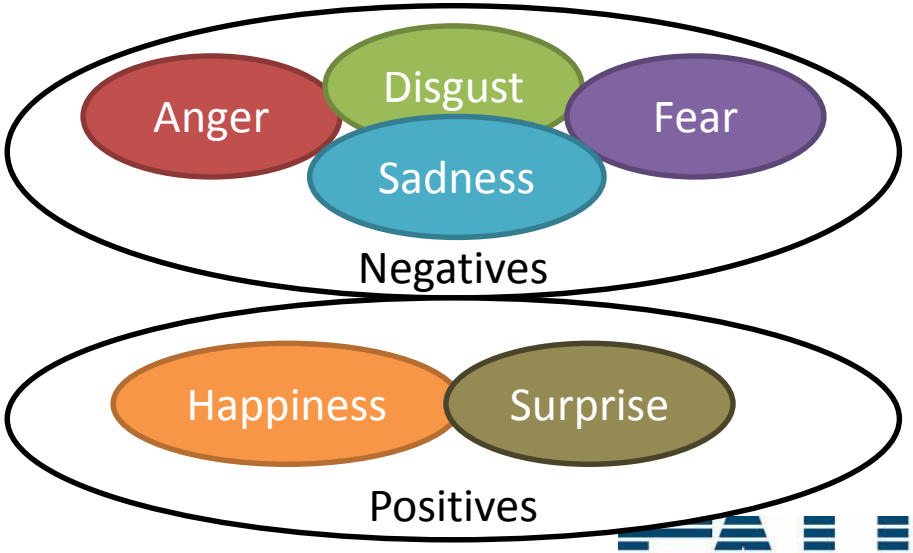
3. Experiments

2-class

Berlin DB



Enterface05 DB



UNIVERSIDAD DE ANTIOQUIA
1803

FAU

Outline

1. Introduction
2. Methodology
3. Databases and experiments
4. **Results**
5. Conclusion

4. Results

Multi-class experiment

Feature Set	Berlin database		Enterface05 database	
	UV	Syllables	UV	Syllables
Feature set 1	62 ± 7	70 ± 9	40 ± 4	41 ± 6
Feature set 2	59 ± 11	78 ± 8	55 ± 5	59 ± 5
Multi-class SVM	65 ± 9	**82 ± 14**	55 ± 5	**60 ± 8**

UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

4. Results

Multi-class experiment

	Berlin Database			Enterface05 Database		
	Fear	Disgust	Anger	Fear	Disgust	Anger
Fear	**80.3**	10.1	8.9	**58.3**	16.2	16.8
Disgust	15.2	**76.7**	8.9	21.4	**59.8**	21.6
Anger	4.5	13.3	**82.2**	20.3	24.0	**61.6**

UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

4. Results

2-class Experiment

Feature Set	Frames	Berlin database			Enterface05 database		
		Acc	Sens	Spec	Acc	Sens	Spec
Feature set 1	UV	80.1	82.9	75.3	55.3	55.5	54.8
Feature set 2	UV	77.8	80.2	70.6	54.0	62.1	37.5
SVM-multi-class	UV	85.6	86.4	78.2	**68.2**	68.2	67.9
Feature set 1	Syll	84.2	85.0	78.2	56.0	60.9	46.2
Feature set 2	Syll	76.3	79.2	67.4	53.9	55.0	31.0
SVM-multi-class	Syll	**86.3**	86.9	81.5	67.9	68.2	63.6

Outline

1. Introduction
2. Methodology
3. Databases and experiments
4. Results
5. **Conclusion**

UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

5. Conclusion

1. New method to characterize non-stationary process in speech.
2. Unvoiced and syllables are characterized.
3. Multi-class and 2-class experimets are considered

jcamilo.vasquez@udea.edu.co

5. Conclusion

4. The method is more suitable to characterize syllables instead of unvoiced frames in multi-class.

5. There is no significative diference in 2-class

6. The second classification provides an improvement in the general accuracy rate relative to the separately classification

UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

5. Conclusion

7. Features proposed could be used as complement to classical features for speech analysis.

8. More features from the TARMA should be analyzed

jcamilo.vasquez@udea.edu.co

Thanks!

UNIVERSIDAD DE ANTIOQUIA
1803

FAU

jcamilo.vasquez@udea.edu.co

Time Dependent ARMA for Automatic Recognition of Fear-Type Emotions in Speech

J.C. Vásquez-Correa[1], J.R Orozco-Arroyave[1,2], J.D Arias-Londoño[1], J.F Vargas-Bonilla[1], L.D. Avendaño[3], Elmar Nöth[2]

[1] Faculty of Engineering, University of Antioquia UdeA
[2] Pattern Recognition Lab, Friedrich Alexander Universität, Erlangen-Nürnberg
[3] Mechanical & Aeronautical Engineering, University of Patras

Text, Speech and Dialogue, TSD 2015