OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

Speaker Recognition

Maik Riestock & Jonas Marquardt

Speaker Recognition

- Identify person by the speech

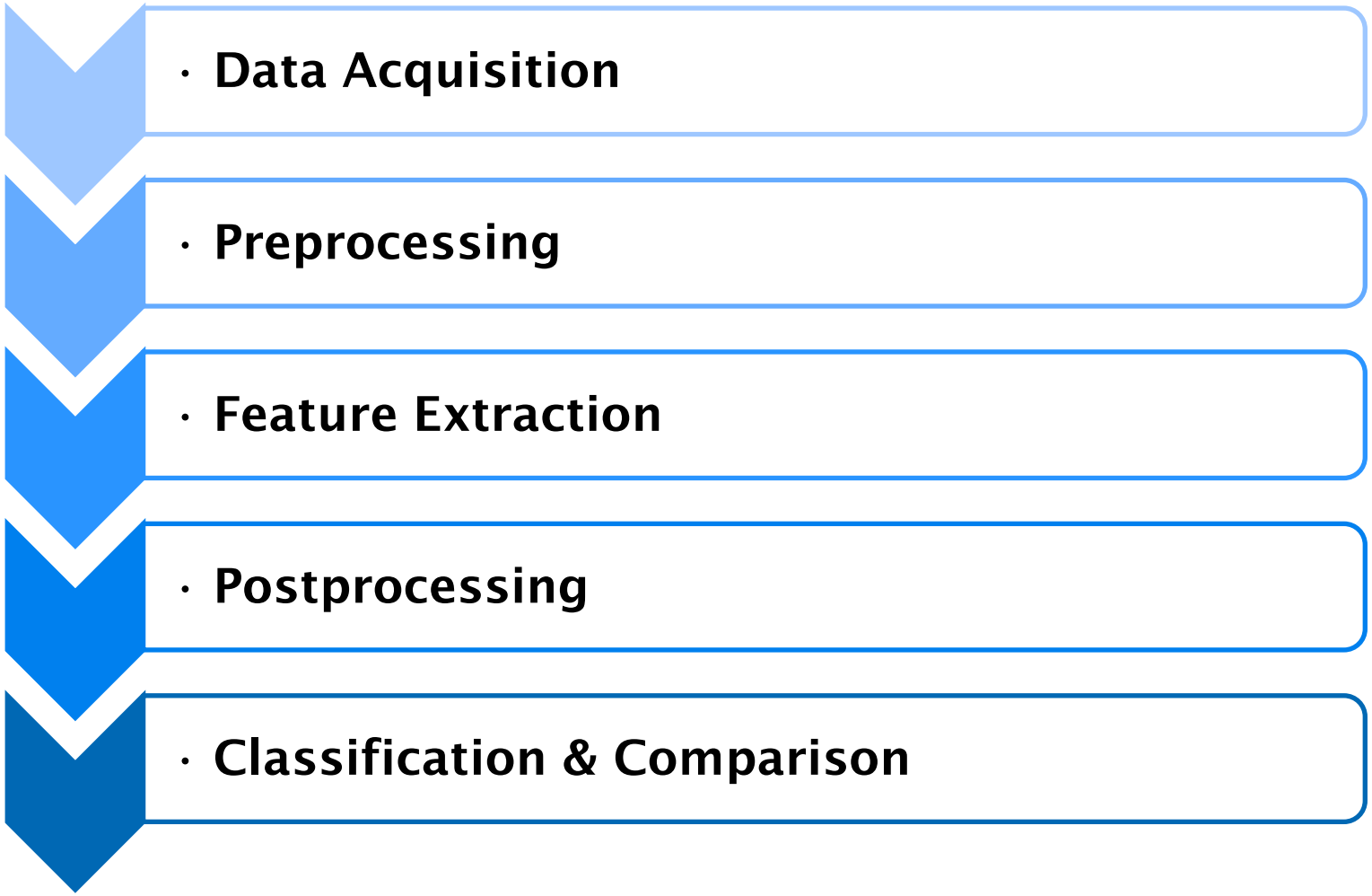
Our task was:

- Closed set speaker authentication on the Hyke speech database
- Compare the results achieved (in terms of authentication performance) to results presented in [Reda2011]
- A projection of the samples in your data set to the characters of ‘Doddingtons Zoo’ [Doddington1998]

Hyke Projekt [Reda2011]

- Von wem?
- Wo?
- Ziel
- Was wurde dafür getan?
- Was hat das mit uns zu tun?

Processing chain



Database from Hyke Projekt[Reda2011]

- Speaker 83:
  - Male: 48
  - Female: 35

- Recordings:
  - mostly 5 recordings of each speaker
  - total recordings: 415
  - length between 5 and 35 seconds
  - overall Size 100mb
  - recorded with normal telephones
  - natural environment

Source: http://www.azreda.org/audio.html

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF | FAKULTÄT FÜR INFORMATIK

Preprocessing

1. Step:

Male/Female

Overall

2. Step:

Train Set

- For training of the classifiers
- 4 recordings each speaker

Test Set

- For evaluation of the classifiers
- 1 recording of 10 seconds each speaker

Feature Extraktion

- Used the AMSL Audio Feature Extractor
  - Divide recordings in small samples
  - Delivers 593 Features each sample

Result:

- All 415 recordings delivered around 18.000 samples

Post processing

- Using Weka
  - Deleted unhelpful Features
    e.g. same value in all samples
  - Deleted samples without speech
    under usage of Amplitude

Approximately 50% of the Database was filtered

(Because it contains no speech)

Classification

- Using Weka
  - Used all relevant classifier
  - Used train set for training of the classifier
  - Used test set for evaluation of the classifier

Best classifiers (correct classified):

	IBK	RandomForest
Male	58,5%	41,8%
Female	54,9%	39,1%
mixed	53,2%	33,9%

Result of classification

- Unterschied samples recordings
- Speaker is correct classified, if the majority of samples of one recording is correct classified
- Tabelle vergleich Hyke vs. unser Ergebnis

Speaker is correct classified, if the majority of samples of **one** recording is correct classified

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF | FAKULTÄT FÜR INFORMATIK

Covarianzmatrix

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae	af	ag	ah	ai	<-- classified as
19	0	0	0	0	1	1	0	0	0	0	0	0	0	2	0	5	3	0	1	0	1	0	1	0	2	0	0	0	0	1	1	0	2	0	a = female_01
1	26	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	1	2	1	1	0	3	0	0	0	3	0	0	0	0	0	b = female_02
0	0	20	0	0	0	2	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	5	1	0	0	0	0	0	0	0	1	0	0	1	c = female_03
0	0	0	8	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	d = female_04
1	3	1	0	12	2	0	0	5	0	0	1	1	0	0	0	0	1	0	4	2	0	0	0	0	1	0	0	0	2	2	0	0	1	1	e = female_05
0	0	0	0	0	36	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	f = female_06
0	1	0	0	2	0	13	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1	3	0	g = female_07
2	0	0	0	0	0	3	26	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	1	0	h = female_08
0	0	0	0	1	0	2	0	11	0	0	0	1	0	3	0	0	1	0	3	0	1	0	2	0	0	0	0	0	2	3	3	0	0	0	i = female_09
0	0	1	0	0	0	0	0	0	22	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	j = female_10
0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	1	0	0	1	0	1	0	1	1	0	2	0	1	0	0	0	0	0	0	1	k = female_11
0	0	0	0	1	0	0	3	0	0	1	17	0	0	0	0	0	0	0	2	0	0	0	1	0	3	0	1	0	0	0	0	0	2	0	l = female_12
0	1	0	0	2	0	0	1	1	0	2	0	16	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	1	1	0	2	0	0	0	m = female_13
1	0	0	4	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	n = female_14
1	1	0	0	6	0	2	0	0	0	0	1	0	0	11	0	0	2	0	0	2	1	1	0	0	4	1	4	0	1	0	1	0	5	0	o = female_15
0	0	0	0	0	0	0	1	0	2	0	0	0	2	0	18	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	0	p = female_16
3	1	0	0	3	0	1	0	1	0	0	0	0	0	2	0	14	5	0	0	0	2	0	3	0	0	0	0	0	1	0	2	1	8	0	q = female_17
3	0	0	0	2	0	0	0	0	0	1	0	1	0	1	0	0	23	0	1	2	0	0	0	0	1	0	0	1	0	0	0	0	2	0	r = female_18
0	0	4	0	0	0	0	0	0	2	4	0	0	1	0	0	0	0	12	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	1	s = female_19
1	3	0	0	1	0	1	1	0	0	1	1	0	0	0	0	1	0	0	15	1	1	0	3	0	0	0	0	0	2	0	3	0	2	0	t = female_20
1	0	0	0	2	1	1	0	0	0	1	1	0	0	1	0	1	1	2	1	8	2	2	0	0	0	0	0	0	0	0	0	0	2	1	u = female_21
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	6	8	1	1	0	1	0	0	1	0	3	2	0	1	1	v = female_22
0	0	0	0	0	0	0	0	0	2	1	1	0	0	1	1	0	0	1	0	1	0	32	1	0	0	0	0	0	0	0	0	0	0	1	w = female_23
1	1	1	0	2	0	0	0	2	0	0	1	0	0	0	0	1	2	1	3	0	0	1	12	0	2	0	1	0	0	0	2	0	2	0	x = female_24
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	17	0	0	0	0	0	0	0	1	0	0	y = female_25
4	5	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	2	0	21	0	0	0	2	1	0	0	1	0	z = female_26
0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2	16	0	0	0	0	0	0	0	0	aa = female_27
1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	0	0	0	0	0	2	0	ab = female_28
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	14	1	0	0	0	0	0	ac = female_29
1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	5	0	0	1	0	0	2	0	1	0	29	1	0	0	0	0	ad = female_30
0	1	0	0	0	0	0	0	1	0	1	0	4	0	1	1	0	0	0	6	0	0	0	2	0	0	0	1	0	0	18	1	0	2	0	ae = female_31
0	1	0	0	1	0	1	0	4	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	5	0	0	1	af = female_32
0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	1	0	2	0	0	1	1	0	0	0	13	0	0	ag = female_33
0	0	0	0	3	0	0	0	0	0	1	1	0	0	2	0	1	1	0	1	3	2	0	0	0	0	0	1	0	0	0	1	0	16	2	ah = female_34
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	13	ai = female_35

Doddington Zoo

- Sheeps: easily accepted by the system
- Goats: exceptionally unsuccessful at being accepted
- Lambs: exceptionally vulnerable to impersonation
- Wolves: exceptionally successful at impersonation

character	percentage
Sheep	90%
Goat	5%
Lamp	2,5%
Wolfs	2,5%

Conclusion

With:

- Non optimal database
- Good feature extraction tool
- Good classification tool
- Minimal post-processing

We accomplish:

- 97,5% authentication rate

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF | FAKULTÄT FÜR INFORMATIK

Sources

[Reda2011] Azarias Reda, Saurabh Panjwani and Edward Cutrell: Hyke: A Low-cost Remote Attendance Tracking
System for Developing Regions. The 5th ACM Workshop on Networked Systems for Developing Regions (NSDR),
2011

[Doddington1998] George Doddington, Walter Liggett, Alvin Martin, Mark Przybocki and Douglas Reynolds:
Sheep, Goats, Lambs and Wolves - A Statistical Analysis of Speaker Performance. In NIST 1998 Speaker
Recognition Evaluation, Int. Conference on Spoken Language Processing, NIST, 1998.

[Vielhauer2006] Claus Vielhauer: Biometric User Authentication for IT Security: From Fundamentals to
Handwriting. Springer, Advances in Information Security, 2006.

http://www.cs.waikato.ac.nz/ml/weka/

http://www.azreda.org/audio.html

Thank you for your attention!

www.ovgu.de