

# Biometrics and Security Speaker Recognition

Jonas Marquardt und Maik Riestock  
Otto von Guericke University Magdeburg

abstract  
abstract  
abstract  
abstract  
abstract

Categories and Subject Descriptors:

Additional Key Words and Phrases: Speaker Recognition, Audio Feature Extraction

## 1. MOTIVATION

Um Personen eindeutig zu identifizieren gibt es verschiedene Verfahren. Eines davon ist die Erkennung der Stimme. Als Aufnahmegerät ist ein handelsübliches Mikrophon ausreichend. In dieser Übung ging es darum, herauszufinden wie eine Stimmenerkennung umgesetzt wird und welche Eigenheiten dieses Verfahren mit sich bringt.

[George Doddington 1998]

## 2. HYKE-SYSTEM

Das Hyke ist ein System welches durch die Stimme eines Kindes stammt von einer dezentralen Bildungseinrichtung aus Indien die es dazu verwendet um nachzuvollziehen ob die Lehrer ihren Unterricht allgemein

- Bildungseinrichtung
- Rajasthan, NW India
- jede Schule hat 1-3 Lehrer
- Hauptquartier in Udaipur, 150 entfernt
- Hauptquartier in Udaipur, 150 entfernt

bisher

- visuelle Kontrolle der Anwesenheit
- 2 mal am Tag Bilder
- manuell jedes Bild verifiziert wurde
- Kosten für Angestellte
- jede Schule benötigt eine digitale Kamera

neu

---

This report was created in the context of the course Biometrics and Security [BIOSEC] winter term 2014/15. This course was held by: Prof. Dr.-Ing. Jana Dittmann and Prof. Dr.-Ing. Claus Viehauer; Research group Multimedia and Security, Otto-von-Guericke-University of Magdeburg, Germany. The course was supported by: Dr.-Ing. Christian Krtzer, M.Sc. Kun Qian

- neuer Ansatz mit Stimmen-Erkennung
- da 75 % der Schulen eh schon ein Telefon haben
- automatische Verifikation

Ergebnis

- 95 % Erkennungsrate, State of the Art

## 3. OUR APPROACH

Aufgabenstellung A

- Closed set speaker authentication on the Hyke speech database
- Compare the results achieved (in terms of authentication performance) to the results presented in [Azarias Reda 2011]
- A projection of the samples in your data set to the characters of 'Doddingtons Zoo'

def. speaker recognition [Beigi 2011] Speaker recognition, sometimes referred to as speaker biometrics, includes identification, verification (authentication), classification, and by extension, segmentation, tracking and detection of speakers. It is a generic term used for any procedure which involves knowledge of the identity of a person based on his/her voice.

Bei unseren Arbeitsschritten haben wir uns dabei an der von ... vorgestellten ProChain gehalten

- Woher? [Prof. Dr. Jana Dittmann 2014]
- genaue Erklärung in den einzelnen Kapiteln

Für diese Aufgabe wurden uns die folgenden Programme bereitgestellt:

- AAFE, ist ein ... und wurde für die Feature Extraction verwendet
- WEKA, ist ein ... und wurde für die RePost und ReClassification verwendet

## 4. DATABASE

- male, female
- recordings, length
- format, compression?
- noise, telephone device

## 5. PRE-PROCESSING

In diesem Kapitel geht es um die Daten unserer Datenbank auf die folgenden Schritte vorzubereiten.

Da wir in unseren Ergebnissen am Ende einen möglichen Unterschied zwischen den Geschlechtern beobachten zu können... dafür wurde das Set mit beiden Geschlechtern, hier mixed set, aufgeteilt in zwei Sets mit ausschließlich Stimmen von weiblichen Sprechern, hier female set, und mit ausschließlich männlichen Sprechern, hier male set.

außerdem benötigen die Klassifikatoren zwei Sets von Daten. Mit dem einen Set wird das Modell trainiert, hier *train set*, und mit dem anderen Set evaluiert, hier *test set*.

- Was haben wir gemacht?
- Warum?
- mit welchem Ergebnis?

eine Aufteilung wie folgt durchgenommen:

- *mixed train set*, beinhaltet ...?
- *mixed test set*
- *female train set*
- *female test set*
- *male train set*

Diese Aufteilung erlaubt es und zu einem die Klassifikatoren mit verschiedenen Sets für Trainings und Evaluierungsphase zu füttern und zum anderen dient uns die Aufteilung in *mixed*, *female* und *male* Splitter einen möglichen Unterschied der Ergebnisse zu beobachten.

## 6. FEATURE EXTRACTION

- Benutzung von AAFE
- Woher kommt er?
- Was macht er?
- Wie?

## 7. POST-PROCESSING

- Benutzung von WEKA
- Was haben wir gemacht?
- Warum?
- mit welchem Ergebnis?
- Befehle aus WEKA

## 8. KLASSIFIKATION

In diesem Kapitel geht es darum Klassifikatoren zu finden die bei der Klassifizierung gut es im Allgemeinen um

Unsere Daten sind dabei in die 6 Sets unterteilt und beinhalten die Features der Recordings.

Zur Bestimmung der besten Klassifizierer haben wir die Methode des *try and error* verwendet. Das heißt wir haben alle möglichen Klassifikatoren in den Default-Setting auf unsere Datenbasis angewandt und danach die Ergebnisse verglichen.

Ein gutes Ergebnis bestand darin dass möglichst viele Samples eines Sprechers dem richtigen Sprecher zugeordnet wurden. Also der Klassifikator unter Verwendung des *mixed test set* eine gute Treffergenauigkeit aufwies.

Dabei hat sich ein Klassifikator als besonders gut erwiesen, der *ibk*. Dieser Klassifikator erzielte ein Ergebnis von 54.94 % Treffergenauigkeit bei dem *mixed test set*. Als Vergleich haben wir den Klassifikator mit dem zweitbesten Ergebnis mit aufgeführt, der *RandomForest*. Die Tabelle I zeigt das Ergebnis beider Klassifikatoren mit den dazugehörigen Konfigurationen des Klassifikators.

Aus den Ergebnissen ist zu entnehmen dass es keinen signifikanten Unterschied zwischen den Ergebnissen der Sets *female test set* und *male test set* existiert. Der bestehende Unterschied lässt sich aus der geringen Größe des Datensets erklären.

Datenset	IBK	RandomForest
female test set	54.94%	39.14%
male test set	58.50%	41.7679%
mixed test set	53.16%	33.86%

Table I.

Ergebnisse der Klassifikation

Außerdem ist zu beobachten dass sich die Ergebnisse bei der Klassifikation verschlechtert hat bei Erhöhung der Anzahl von Sprechern. Dies ist jedoch ein zu erwartendes Ergebnis da der Klassifikator nun das Sample eines Sprechers mit 82 anderen Samples statt mit 47 bzw. 34 anderen Samples vergleichen muss.

### 8.1 Authentifizierung

In diesem Kapitel geht es nun um die Aufgabe der *Closed set speaker authentication*. Wobei die akustische Aufnahme eines Sprechers mit der aller anderen möglichen Sprechern verglichen wird und die beste Übereinstimmung als Ergebnis ausgegeben wird. [Beigi 2011] Zu beachten ist, dass hier im Gegensatz zu der *Open set speaker authentication* es in jedem Fall zu einem Ergebnis kommt.

Die Aufgabe besteht nun darin das Ergebnis der Klassifikation der Samples zu interpretieren. Dafür betrachten wir jeden Sprecher die Verteilung seiner Samples. Hierbei wurde ein Sprecher richtig erkannt wenn bei ihm die größte Menge Samples zugeordnet wurden. Dies bedeutet dass wir auch mit einer geringen Anzahl richtig klassifizierter Samples einen Sprecher authentifizieren konnten solange die übrigen Samples gleichmäßig verteilt waren.

Dieses Verfahren wurde bei allen Sprechern angewandt und das entstandene Ergebnis ist zu sehen in Tabelle II.

Datenset	Gesamt	Richtig	Falsch	Anteil
female test set	35	33	2	94.29%
male test set	48	46	2	95.83%
mixed test set	83	79	4	95.18%

Table II.

Ergebnisse der Authentifizierung

Das Ergebnis von 95.18% richtig erkannten Sprechern ist gut und entspricht damit dem *state-of-the-art*. [Beigi 2011] Im Vergleich dazu wurde im *Hyke*-Projekt ein Ergebnis von 95% erreicht, welches mit unseren nahezu identisch ist. [Azarias Reda 2011]

## 9. DODDINGTONS ZOO

Doddingtons Zoo geht es darum dass Sprecher ein unterschiedliches Verhalten bezüglich den Erfolg ihrer Authentifikation aufzeigen. [George Doddington 1998] dabei lassen sich Sprecher in

4 kategorien unterscheiden die jeweils von einem tier repräsentiert werden.

Eine Beschreibung dieser Kategorien:

- Sheeps:** auerordentlich *leicht* von dem System erkannt
- Goats:** auerordentlich *schwer* von dem System erkannt
- Lambs:** auerordentlich *verwundbar* gegenber Nachahmung
- Wolves:** auerordentlich *erfolgreich* bei der Nachahmung anderen Sprecher

[Prof. Dr. Jana Dittmann 2014]

Animal	female	male	mixed	Anteil
Sheep	32	43	75	90.36%
Goat	2	2	4	4.82%
Lamb	1	1	1	2.41%
Wolf	0	2	2	2.41%

Table III.  
Ergeb-  
nisse  
der  
Klas-  
si-  
fika-  
tion

interpretation

- Vorstellung des Zoos
- Anwendung
- Results

## 10. CONCLUSION

- past stuff

## 11. FUTURE WORK

- future stuff

## REFERENCES

- Edward Cutrell Azarias Reda, Saurabh Panjwani. 2011. Hyke: A Low-cost Remote Attendance Tracking System for Developing Regions. *Networked System for Developing Regions* (2011).
- Homayoon Beigi. 2011. *Fundamentals of Speaker Recognition*. Springer Science+Business Media.
- Alvin Martin-Mark Przyboc Douglas Reynolds George Doddington, Walter Liggett. 1998. SHEEP, GOATS, LAMBS and WOLVES - A Statistical Analysis of Speaker Performance in the NIST 1998 Skeaper Recognition Evaluation. *National Institute of Standards and Technology* (1998).
- Prof. Dr.-Ing. Claus Vielhauer Prof. Dr. Jana Dittmann. 2014. *Biometrics and Security - Lecture*. Faculty of Computer Science, Institute of Technical and Business Information Systems, Advanced Multimedia and Security Lab (AMSL).

## List of Tables

## APPENDIX

### A. TASK DESCRIPTION

Run your prototype on the collected data and perform a performance evaluation with your prototype. The evaluation must include:

The evaluation must include:

- Closed set speaker authentication on the Hyke speech database
- Compare the results achieved (in terms of authentication performance) to th results persented in
- A projection of the samples in your data set to the characters of 'Doddingtons Zoo'