

# Digitale Signalverarbeitung und Mustererkennung (Digitale Signalverarbeitung 2)

## Abschlussprojekt zur Vorlesung

**Dozent:** Prof. Dr. Volker Stahl

**Autoren:** Andreas Schneider 198805  
Gustav Willig 197332  
Lisa-Franziska Schäfer 199318

# Inhaltsverzeichnis

<b>1</b>	<b>Aufgabenstellung</b>	<b>2</b>
<b>2</b>	<b>Planung und Vorbereitungen</b>	<b>3</b>
<b>3</b>	<b>Klassifikation</b>	<b>4</b>
3.1	Einzelwörterkennung . . . . .	4
3.2	Wortfolgenerkennung . . . . .	5
<b>4</b>	<b>GUI</b>	<b>7</b>

# 1 Aufgabenstellung

Im Rahmen der Mastervorlesung „*Digitale Signalverarbeitung 2*“ bzw. „*Digitale Signalverarbeitung und Mustererkennung*“ soll ein Abschlussprojekt bearbeitet werden, um die aus der Vorlesung gewonnenen theoretischen Erkenntnisse praktisch anzuwenden. Ziel des Projekts ist es ein Programm zur einfachen Spracherkennung zu realisieren. Zudem sollen am Computer einige Experimente durchgeführt und in dem vorliegenden Abschlussprojektbericht dokumentiert werden.

## 2 Planung und Vorbereitungen

Die Bearbeitung der Projektaufgabe erfolgt in der Programmiersprache „*Python*“, da die Gruppenmitglieder hier bereits einige Vorkenntnisse mitbringen. Ein weiterer Vorteil von Python ist dessen Plattformunabhängigkeit, wodurch das Projekt ohne Anpassungen sowohl auf Linux- als auch auf Windows-Betriebssystemen lauffähig ist.

Zum Trainieren von verschiedenen Modellen wurden zunächst die Wörter in der nachfolgenden Tabelle 1 erzeugt. Für jedes Wort liegen insgesamt 54 Merkmalsvektorfolgen vor. Dabei stammen jeweils 36 von zwei männlichen Sprechern und 18 von einer weiblichen Sprecherin. Um die Abhängigkeit der Klassifikationsqualität von der Umgebung zu vermindern, wurden die Aufnahmen zur Hälfte in einer ruhigen Umgebung (G117) und mit Hintergrundgeräuschen (A409, laufender Beamer) gesprochen. Weiterhin wurde eine Aufnahme der im jeweiligen Raum herrschenden „*Stille*“ erzeugt.

Namen	Andreas, Gustav, Lisa
Zahlen	Zwei, Drei
Hausaufgaben	Signalverarbeitung, Fouriertransformation
sonstige	mögen, und

**Tabelle 1:** Erzeugte Merkmalsvektorfolgen

### 3 Klassifikation

Da die Software im Laufe des Projektes stetig weiterentwickelt wurde, steht zur Berechnung der Kosten bzw. Wahrscheinlichkeiten nur die am Besten funktionierende Methode zur Verfügung, „HMM mit maximum Approximation“.

**Notiz:** Bei Einzelwörtern hat der Sprecher ein besseres Ergebnis als All. Bei Wortfolgen schneidet das Modell All besser ab, als der eigentliche Sprecher!

#### 3.1 Einzelworterkennung

Die folgenden Tabellen zeigen ...

Stillemodell	Referenzmodell	Getestetes Wort	Kosten
A409	Drei_Andreas_A409	Drei_Andreas	-168,04
		Drei_Gustav	198,98
		Drei_Lisa	113,26
A409	Drei_All_409	Drei_Andreas	-95,74
		Drei_Gustav	-46,10
		Drei_Lisa	-112,95
G117	Drei_Andreas_A409	Drei_Andreas	-179,47
		Drei_Gustav	89,10
		Drei_Lisa	12,85
G117	Drei_All_A409	Drei_Andreas	-105,82
		Drei_Gustav	-57,20
		Drei_Lisa	-125,85

**Tabelle 2:** Getestete Prädiktion für das Stillemodell A409

Stillemodell	Getestet für	Wort	Sprecher	Kosten
A409	Signalverarbeitung_Gustav_G117	Signalverarbeitung	Gustav	230,60
G117	Signalverarbeitung_Gustav_G117	Signalverarbeitung	Gustav	225,27
A409	Zwei_Lisa_A409	Zwei	Lisa	-210,28
G117	Zwei_Lisa_A409	Zwei	Lisa	-235,00

**Tabelle 3:** Prädiktion mit Sprechererkennung

Stillemodell	Getestet für	Wort	Kosten
A409	Signalverarbeitung_Gustav_G117	Signalverarbeitung	447,73
G117	Signalverarbeitung_Gustav_G117	Signalverarbeitung	442,40
A409	Zwei_Lisa_A409	Zwei	28,13
G117	Zwei_Lisa_A409	Zwei	21,17

Tabelle 4: Prädiktion mit veralgemeinerten Modellen

### 3.2 Wortfolgenerkennung

Da es wie bei der Einzelworterkennung kaum einen Unterschied zwischen dem verwendeten Stillemodell gibt, werden für die folgenden Tests nur G117 verwendet. Aufgrund der sonst übergroßen Breite der Tabellen wird das Wort „*Fouriertransformation*“ mit FT und „*Signalverarbeitung*“ mit SV abgekürzt.

Gesprochenes Wort	Andreas	Gustav	Lisa
Lisa	Lisa	Lisa	Lisa
und	und	und	und
Gustav	Gustav	Gustav	Gustav
und		und	und
Andreas	Andreas	Andreas	Andreas
mögen	mögen	mögen	mögen
SV	SV	SV	SV

Tabelle 5: Prädiktion mit veralgemeinerten Modellen

Gesprochenes Wort	Andreas		Gustav		Lisa	
	Wort	Sprecher	Wort	Sprecher	Wort	Sprecher
Lisa	Lisa	Andreas	Lisa	Gustav	Lisa	Lisa
und	und	Andreas	und	Gustav	und	Lisa
Gustav	Gustav	Andreas	Gustav	Gustav	Gustav	Lisa
und			und	Gustav	und	Lisa
Andreas	Andreas	Andreas	Andreas	Andreas	Andreas	Andreas
mögen	mögen	Gustav	mögen	Gustav	mögen	Lisa
SV	SV	Andreas	SV	Gustav	SV	Lisa

Tabelle 6: Prädiktion mit Sprechererkennung

Gesprochenes Wort	Andreas	Gustav	Lisa
		moegen	drei
und	und	und	und
und	und	und	und
		drei	
FT	FT	FT	FT
zwei	zwei	zwei	zwei
drei	drei	drei	drei

**Tabelle 7:** Prädiktion mit veralgemeinerten Modellen

Gesprochenes Wort	Andreas		Gustav		Lisa	
	Wort	Sprecher	Wort	Sprecher	Wort	Sprecher
	und	Andreas	drei	Andreas	zwei	Andreas
und	und	Gustav	und	Gustav	und	Lisa
und	und	Andreas	und	Gustav	und	Lisa
			drei	Andreas		
FT	FT	Andreas	FT	Gustav	Andreas	Andreas
zwei	drei	Andreas	drei	Andreas	zwei	Lisa
drei	drei	Andreas	drei	Gustav	drei	Lisa

**Tabelle 8:** Prädiktion mit Sprechererkennung

**Notiz:** je größer die Pause zwischen den einzelnen Worten desto besser das Klassifikationsergebnis

## 4 GUI

Zur besseren Bedienbarkeit des Programms wurde eine graphische Benutzeroberfläche entwickelt. Diese ermöglicht es mit geringem Aufwand, Modelle aus den vorhandenen Merkmalsvektorfolgen zu trainieren sowie zu testen. Hierdurch wird der Umgang mit den Datenmengen sowie die Durchführung von Experimenten vereinfacht. Die Benutzeroberfläche ist in der Abbildung 1 dargestellt.

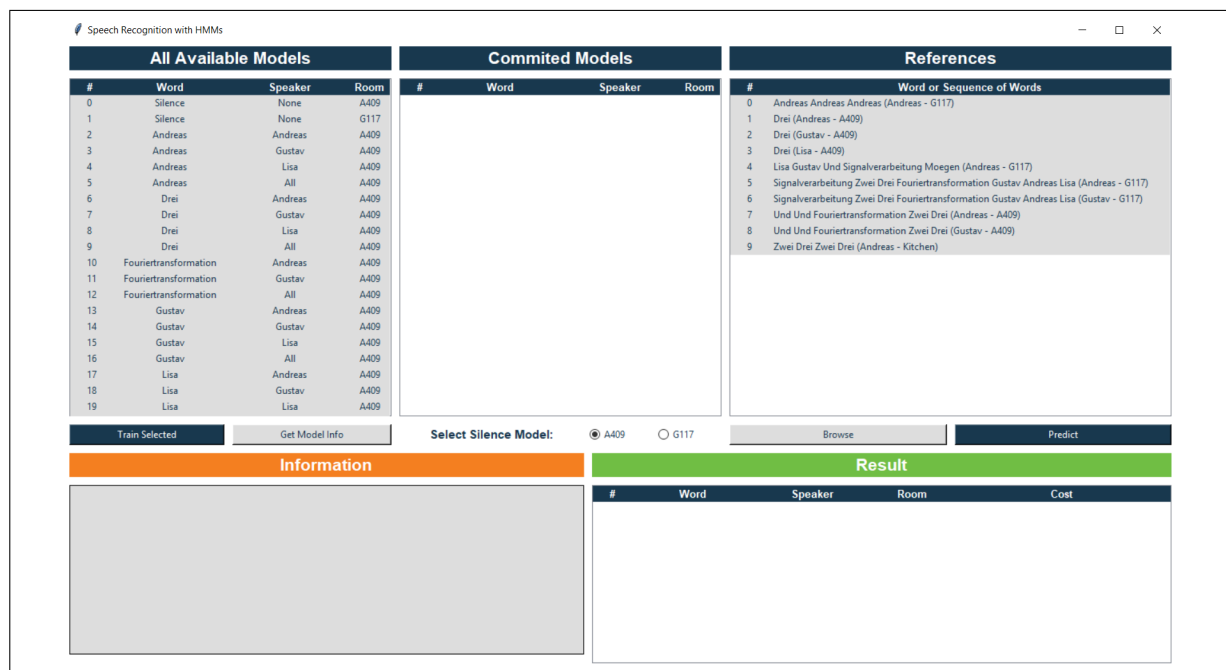


Abbildung 1: grafische Benutzeroberfläche

In der linken oberen Box sind alle verfügbaren Modelle aufgelistet. Über die beiden darunter liegenden Buttons können die Modelle entweder neu trainiert oder Informationen, wie z.B. die Modelllänge oder die benötigten Iterationen, angezeigt werden. Die mittlere Box zeigt die aktuell für die Klassifikation zur Auswahl stehenden Modelle an. Durch einen Doppelklick auf ein Modell in der linken Box, kann dieses für die später folgende Prädiktion geladen werden. Das zu verwendende Stillemodell wird über die darunter angeordneten Radiobuttons ausgewählt.

Um ein Wort durch die Software klassifizieren zu lassen, muss dieses durch den Button „Browse“ in die obere rechte Box geladen und ausgewählt werden. Hierdurch können auch gesprochene Wortfolgen ausgewertet werden. Zum Starten der Prädiktion wird der Button „Predict“ betätigt, das Ergebnis wird danach rechts unten angezeigt. Hierbei wird sowohl das erkannte Wort, als auch der ermittelte Sprecher und die benötigten Kosten aufgelistet.