Karlsruhe Institute of Technology

# Signalinterpolation für Elektrodenarrays in der EMG-basierten Spracherkennung

Bachelorarbeit am Cognitive Systems Lab
Prof. Dr.-Ing. Tanja Schultz
Fakultät für Informatik
Karlsruher Institut für Technologie

von

**Davud Adigüzel**

Betreuer:
Dipl-Math. Michael Wand
Prof. Dr.-Ing. Tanja Schultz

Tag der Anmeldung: 03. 12. 2012
Tag der Abgabe: 02. 04. 2013

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, den 01. 02. 2013

**Zusammenfassung**

Spracherkennung bezeichnet das Ableiten von Sprache aus vom Menschen bzw. menschlichen Körper ausgesandten Biosignalen, die an der Spracherzeugung beteiligt sind. Das sind unter anderem akustische, visuelle, aber auch elektrische Signale. Letztere sind vor allem dann interessant, wenn der Gebrauch von akustischer Sprache nicht gewollt oder nicht möglich ist.

Mittels Elektroden können die elektrischen Signale der Muskelgruppen im Gesicht, die an der Spracherzeugung beteiligt sind, erfasst und in komplexen Algorithmen die ursprüngliche Sprachinformation extrahiert werden. Allerdings gilt es auf dem Weg zu einem alltagstauglichen Spracherkennungssystem viele Hürden zu überwinden. Zum einen wirkt das Gewebe, das elektrische Signale passieren müssen, bevor sie eine Elektrode erreichen, dämpfend, zum andren überlagern sich Signale gegenseitig. Außerdem treten eine große Zahl von Störeinflüssen auf, die die Spracherkennung erschweren.

Als ein Schritt in Richtung robuste Spracherkennung werden hier am Cognitive Systems Labs seit einigen Jahren Elektrodenarrays erforscht. Durch das Sammeln von redundanten Informationen, sollen Fehler ausgeglichen werden können. An dieser Stelle setzt die in dieser Arbeit entwickelte Signalinterpolation an. Es wurde herausgefunden, dass Störungen in einzelnen Kanälen durch Informationen aus benachbarten Kanälen entgegengewirkt werden kann, ohne dabei das Messergebnis gravierend zu verschlechtern.

# Inhaltsverzeichnis


**1 Einleitung** **1**
1.1 Nutzen von Silent Speach Interfaces . . . . . 1
1.2 Zielsetzung der Arbeit . . . . . 3
1.3 Gliederung der Arbeit . . . . . 3

**2 Grundlagen** **5**
2.1 Entstehung des EMG-Signals, Muskelkontraktion . . . . . 5
2.2 Messung des EMG-Signals . . . . . 6
2.3 Das EMG-basierte Spracherkennungssystem . . . . . 7

**3 Durchführung des Experiments** **9**
3.1 Verwendeter Korpus . . . . . 9
3.2 Beschreibung der Aufnahmesoftware . . . . . 10
3.3 Beschreibung der Hardware . . . . . 11
3.4 Vorverarbeitung in mit/ohne PCA . . . . . 12
3.4.1 Berechnung der Time Domain Features . . . . . 13
3.4.2 Durchführung des LDA-Algorithmus . . . . . 13
3.4.3 Durchführung des PCA-Algorithmus . . . . . 13
3.5 Erkenner . . . . . 14
3.5.1 Janus . . . . . 14
3.5.2 BioKIT . . . . . 15

**4 Experimente und Resultate** **17**
4.1 Signalinterpolation . . . . . 17
4.2 Auswertung durch Spracherkennung . . . . . 20
4.2.1 Phonembasierte Auswertung mit BioKIT . . . . . 20
4.2.2 BDPF-basierte Auswertung mit Janus . . . . . 20
4.2.3 Vergleich Janus/BioKIT . . . . . 21
4.3 Analyse . . . . . 22
4.3.1 Beste Methode . . . . . 22
4.3.2 Besonderheiten . . . . . 24
4.3.3 Anwendungsszenario . . . . . 25

**5 Zusammenfassung und Ausblick** **27**

**A Anhang** **29**

**Literaturverzeichnis** **45**

# 1. Einleitung

Dieses Kapitel soll dem Leser einen Einblick in die Zielsetzung und Gliederung dieser Arbeit verschaffen. Als Motivation wird es mit einer Einführung in Systeme zur Verarbeitung von lautloser Sprache, die im Weiteren nur noch als Silent Speach Interfaces (kurz: SSI) aufgeführt werden, beginnen.

## 1.1 Nutzen von Silent Speach Interfaces

Schon seit den sechziger Jahren ist die maschinelle Verarbeitung von lautloser Sprache Forschungsthema und findet Anwendung in Umgebungen, in denen akustische Kommunikation z.B. durch Lärm oder Ruhevorschriften eingeschränkt ist. Ein Handy, das über ein SSI verfügt, könnte Privatheit sogar an öffentlichen Plätzen gewährleisten und gerade für Menschen, die aufgrund einer körperlichen Behinderung in ihren Kommunikationsmöglichkeiten eingeschränkt sind, können SSIs einen Weg über die Sprachbarriere bieten. Zur Umsetzung eines SSIs finden sich in der Literatur vielfältige Ansätze:

Da Sprache erkannt werden soll, ist ein naheliegender und direkter Ansatz den Vokaltrakt, der an der Spracherzeugung maßgeblich beteiligt ist, zu untersuchen. In der elektromagnetischen Artikulographie(siehe [4]) erzeugen Senderspulen ein elektromagnetisches Wechselfeld, das Signale in Sensoren induziert, die z.B. auf der Zunge angebracht sind. Aus der Intensität der Signale können dann die Abstände der Sensoren von den Sendern abgeleitet werden. Die Vorteile dieser Methode sind die hohe räumliche und zeitliche Auflösung, die Daten sind im Vergleich zu bildgebenden Verfahren, die z.B. Ultraschallbilder oder gewöhnliche Videobilder liefern, leicht zu verarbeiten und die Methode ermöglicht die Untersuchung der Koordinaten von Zunge, Kiefer und Lippen beim Sprechen. Allerdings kann auf diese Weise die Zunge nur diskret abgetastet und in ihrer kontinuierlichen Form nicht erfasst werden. Ein weiteres Problem sind die relativ teuren magnetischen Sensoren, die direkt an Zunge, Kiefern und Lippen angebracht den Komfort des Probanden beeinträchtigen.

Eine weitere Methode den Vokaltrakt direkt zu untersuchen ist ihn mittels Ultraschall abzutasten (z.B. [5]). Im 2006 veröffentlichten Projekt Ouisper[6] wurden die

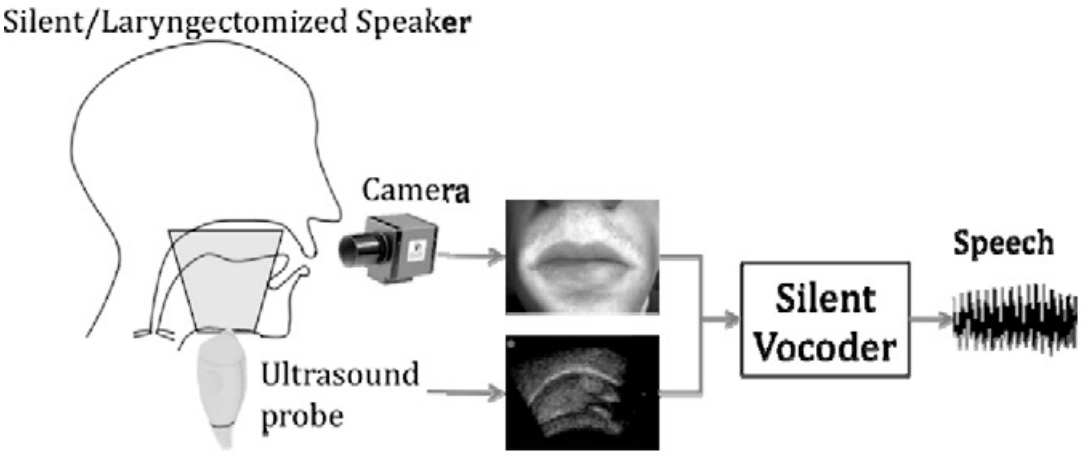


Abbildung 1.1: Schematischer Aufbau eines Spracherkennungssystem basierend auf Ultraschallbildern des Vokaltrakts und Videoaufnahmen der Lippenbewegung [2]

so gewonnen Ultraschallbilder mit Daten einer herkömmlichen Videokamera gekoppelt, die die Lippenbewegungen des Probanden aufnahm. Abbildung 1.1 stellt den Aufbau des Verfahrens schematisch dar.

Natürlich kann auch in eingeschränkter Form das Sprachsignal selbst verwendet werden. Unter "nicht hörbares Murmeln"(Non audible murmur, NAM, siehe [2]) fallen Geräusche, die durch den Luftstrom im Kehlkopf gebildet werden und so leise sind, dass sie von nahe stehenden Personen nicht oder nur kaum gehört werden können. Diese Geräusche können mittels eines hochsensitiven Mikrofons an der Hautoberfläche in der Mundregion aufgenommen werden. Dabei ist es möglich das Mikrofon so zu konfigurieren, dass es Umgebungsgeräusche völlig ausblendet.

Ähnlich wie bei NAM kann das Sprachsignal am Hals in der Nähe des Kehlkopfes oder direkt an Schädelknochen z.B. unterhalb des Ohrs aufgenommen werden. Anstelle von Luft dient der menschliche Körper bzw. Knochen als Übertragungsmedium. Schließlich wird das Signal von einem Hals- oder Knochenmikrofon gemessen, das nur die Vibration des Sprachsignals, nicht aber Umgebungsgeräusche aufnimmt. Darum ist dieser Ansatz besonders in Umgebungen mit hohem Lärmpegel interessant und wurde seit 2000 z.B. im DARPA Advanced Speech Encoding Programm des United States Defense Departments für militärische Zwecke erforscht.

In Fällen in denen kein akustisches Signal zur Verfügung steht sind Methoden sinnvoll, die elektromagnetische Potentiale messen, die im Zusammenhang mit Spracherzeugung entstehen. Die Elektroglottographie beschäftigt sich mit elektromagnetischen Potentialen, die beim Öffnen und Schließen der Stimmlippen entstehen. Hierzu werden Elektroden an beiden Seiten des Kehlkopfes angebracht.

In der Elektromyographie (EMG, [12]) nehmen Oberflächenelektroden elektrische Signale auf, die entstehen, wenn Muskeln aktiviert werden. Die Elektroden werden so im Gesicht und am Hals platziert, dass sie die elektromagnetischen Potentiale erfassen, die beim Ansteuern der an der Spracherzeugung beteiligten Muskelgruppen erfassen können. Da das erfasste Signal vollständig auf Muskelaktivität beruht, kann der akustische Teil der Spracherzeugung weggelassen werden. So kann ein EMG-basiertes Spracherkennungssystem lautlose Sprache aus bloßen Muskelbewegungen

erkennen und ist damit interessant für Menschen, die z.B. aufgrund einer körperlichen Behinderung die Stimmbänder nicht mehr zur Spracherzeugung einsetzen, aber noch die Geschichtsmuskeln bewegen können.

Zuletzt sei noch die Enzephalographie (EEG) erwähnt. Hier messen Elektroden, die in einer Kappe direkt am Cortex angebracht werden, elektrische Potentiale die z.B. durch Reize, denen der Proband ausgesetzt ist, oder durch dessen Gedanken beeinflusst werden können. M.Wester [13] beschäftigte sich 2006 mit der Frage, ob aus EEG-Daten Sprache rekonstruiert werden können. Besonders für Locked-In-Patienten wäre ein solcher Ansatz eines SSIs interessant, da keine Muskelaktivität für einen EEG-Spracherkenner nötig wäre. Wann der erste zuverlässige EEG-basierte Spracherkenner entwickelt werden wird, ist allerdings noch offen.

Eine ausführliche Übersicht und Bewertung über die einzelnen Ansätze für SSI bietet denby et all. 2010 [2]

## 1.2 Zielsetzung der Arbeit

Das Ziel dieser Arbeit war es, die Robustheit des am Cognitive System Labs entwickelten Aufnahmesystems für EMG-basierte Spracherkennung zu erhöhen. Seit 2011 kommen hier Elektrodenarrays zum Einsatz, die die bisher verwendeten einzeln im Gesicht angebrachten Elektroden ersetzen. Elektrodenarrays wie einzeln angebrachte Elektroden unterliegen Störeinflüssen wie etwa dem 50-Herz-Brummen von Steckdosen oder Signalen von Muskelbewegungen, die von der Spracherzeugung unabhängig sind.

Die Fragestellung dieser Arbeit war es, herauszufinden, ob einzelne Kanäle innerhalb eines Elektrodenarrays durch die interpolierten Signale der umliegenden Kanäle ersetzt werden können. So könnten Störungen, die einzelne oder nur sehr wenige Kanäle betreffen, beseitigt werden. Grund zur Annahme, dass auf diese Weise die Robustheit des Verfahrens erhöht werden kann, liefert die Beobachtung, dass Signale benachbarter Kanäle kaum Unterschiede aufweisen.

Bestandteil dieser Arbeit war es außerdem verschiedene Interpolationsmethoden auf unterschiedlichen Erkennern und Aufnahmesessions zu testen und deren Verhalten zu untersuchen.

Das Ergebnis der Arbeit war, dass gezielte Interpolation in Einzelfällen sogar die Fehlerrate herabsetzen kann und im allgemeinen keine deutliche Verschlechterung der Erkennung bedeutet.

## 1.3 Gliederung der Arbeit

Nach der Einleitung führt diese Arbeit in die Grundlagen 2 der EMG-basierten Sprachverarbeitung ein. Es soll dargestellt werden, wie ein EMG-Signal im Körper entsteht, wie es mit Hilfe von Elektroden gemessen werden kann und wie Spracherkennung daraus das ursprünglich Gesagte rekonstruiert. Diese Arbeit baut auf ein bestehendes System zur Spracherkennung auf, dessen Bestandteile in Software und Hardware 3 ebenfalls beschrieben werden sollen. Den Haupteil 4 der Arbeit bildet der Algorithmus zur Durchführung von gewichteter Signalinterpolation und dessen Auswertung unter Verwendung von Spracherkennern in Janus und Biokit. Schließlich folgt eine Zusammenfassng 5 der Arbeit und ein Ausblick auf mögliche Verwendung der hier vorgestellten Verfahren.

# 2. Grundlagen

In diesem Kapitel sollen die Grundlagen der EMG-basierten Spracherkennung dargestellt werden. Dabei wird zuerst darauf eingegangen, wie Muskelkontraktion zu messbaren Depolarisationen führt. Diese wiederum können von Elektroden erfasst und von Spracherkennern verarbeitet werden, so dass schon die Bewegung der an der Spracherzeugung beteiligten Gesichtsmuskeln ausreicht, um daraus Sprache rekonstruieren zu können.

Auf dieser Basis wird der weitere Verlauf dieser Arbeit darauf zielen, Möglichkeiten zu untersuchen, um die Robustheit eines EMG-basiertes Spracherkennungssystems, das Elektrodenarrays verwendet, zu erhöhen.

## 2.1 Entstehung des EMG-Signals, Muskelkontraktion

Überall im menschlichen Körper laufen ständig chemische Reaktionen ab. Zellen kommunizieren untereinander, tauschen Botenstoffe aus, Ladungen fließen. Ist die Ladungsverschiebung ausreichend groß, kann sie an der Hautoberfläche von Elektroden gemessen werden. Insbesondere ist das bei Muskelaktivität der Fall.

Die menschliche Muskulatur lässt sich anhand ihrer histologischen Struktur bzw. der Art der Kontraktion in zwei Untergruppen unterscheiden: Der glatten Muskulatur, die Bestandteil der Inneren Organe ist und die vom zentralen Nervensystem gesteuert autonom ist, und der quer gestreiften Muskulatur oder Skelletmuskulatur, die den willentlich steuerbaren Teil der Muskulatur ausmacht und dem Menschen Bewegungen ermöglicht.

Da die Muskeln, die an der Spracherzeugung beteiligt sind, quer gestreift sind, ist die glatte Muskulatur für diese Arbeit nicht weiter interessant.

Skelletmuskeln setzen sich aus Muskelfasern zusammen, die ihrerseits aus Myosinfilamenten und Aktinfilamenten bestehen. Diese beiden fadenartigen Filamente sind

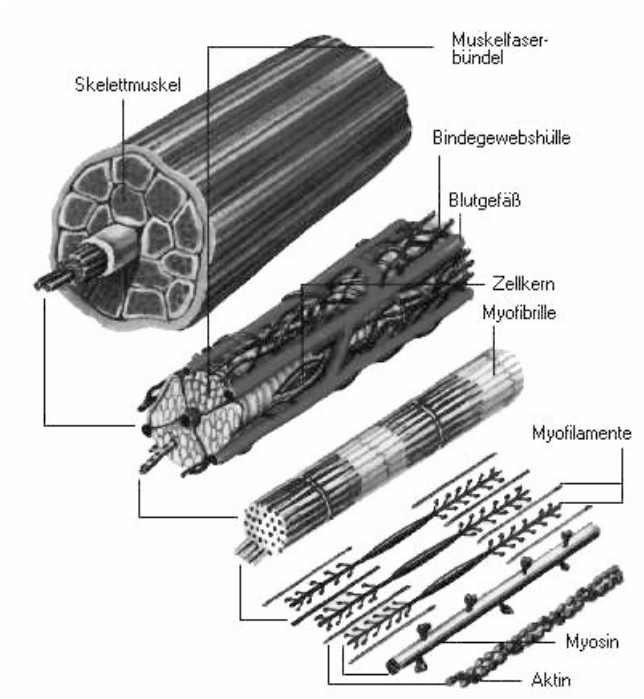


Abbildung 2.1: Aufbau eines Skelettmuskels [1]

parallel gelagert und berühren sich an den Stellen, an denen die Myosinköpfe die Aktinfilamente berühren. Der Aufbau eines Skelletmuskels ist in Abbildung 2.1 dargestellt. Wird die Muskelfaser aktiviert verschieben sich Myosin- und Aktinfilamente unter Energieaufwand gegeneinander und der Muskel wird verkürzt. Dieser Vorgang wird auch Muskelkontraktion genannt und macht die eigentliche Bewegung des Menschen aus.

Der Kontraktionsprozess [7] wird durch das Aktivieren eines Alpha-Motoneurons durch das zentrale Nervensystem eingeleitet. Das Axon des Aplha-Motoneurons verläuft über das Rückenmark und die Peripherie und endet in einer oder mehreren Synapsen des Muskels, die auch motorische Endplatte genannt werden. Feuert das Alpha-Motoneuron, wird die Endplatte aktiviert, die ein Aktionspotential auslöst. Dieses breitet sich von der Endplatte längs der Muskelfaser aus und löst die Kontraktion aus. Je nach Anzahl der Muskelfasern, die mit der Endplatte verbunden sind und von einem Alpha-Motoneuron angesteuert werden, handelt es sich dann um eine großräumige Bewegung, etwa der Beine oder Arme, und es kommen viele Muskelfasern also viel Kraft zum Einsatz oder es handelt sich um feinere Bewegungen wie etwa die Mimik oder Spracherzeugung, so das ein Motoneuron wenige Muskelfasern gezielt ansteuert.

## 2.2 Messung des EMG-Signals

Während der Kontraktion ändern sich die Potentialdifferenzen der angesteuerten Zellen. Zwischen Zellinnerem und Außenraum besteht ein Ruhepotential bei ca. -80 bis -90 mV, das bei Erregung bis ca. 30 mV ansteigt. Dieser Vorgang heißt auch Depolarisation und wird von einem sofortigen Repolarisationszyklus gefolgt,

in dem das Potential in seinen ursprünglichen Zustand zurückfällt. Ist die Zahl der erregten Zellen ausreichend hoch, können die daraus resultierenden Potentialveränderungen an der Hautoberfläche von Elektroden erfasst werden. Alternativ können auch Nadelelektroden genutzt werden, die bis in die Zelle eindringen und so ein räumlich genaueres Signal liefern. Diese Elektroden erfassen auch kleine und tiefgelegene Muskeln und sind gegen Störsignalen von anderen im Körper entstehenden Potentialen robuster als Oberflächenelektroden. Der Nachteil von Nadelelektroden ist, dass sie nur unter klinischen Bedingungen eingesetzt werden können und nur ein Arzt sie anbringen darf. Deshalb verwenden wir für EMG-basierte Spracherkennung ausschließlich Oberflächenelektroden, die zwar ein durch Muskel und Gewebe gefiltertes Signal liefern, dafür aber risikofrei und ohne invasiven Eingriff eingesetzt werden können. In Folge dieser Entscheidung hängt das EMG-Signal stark von der Positionierung der Elektroden ab. Diese richtet sich nach an der Spracherzeugung beteiligten Muskeln, die in Abbildung 2.2 dargestellt sind.

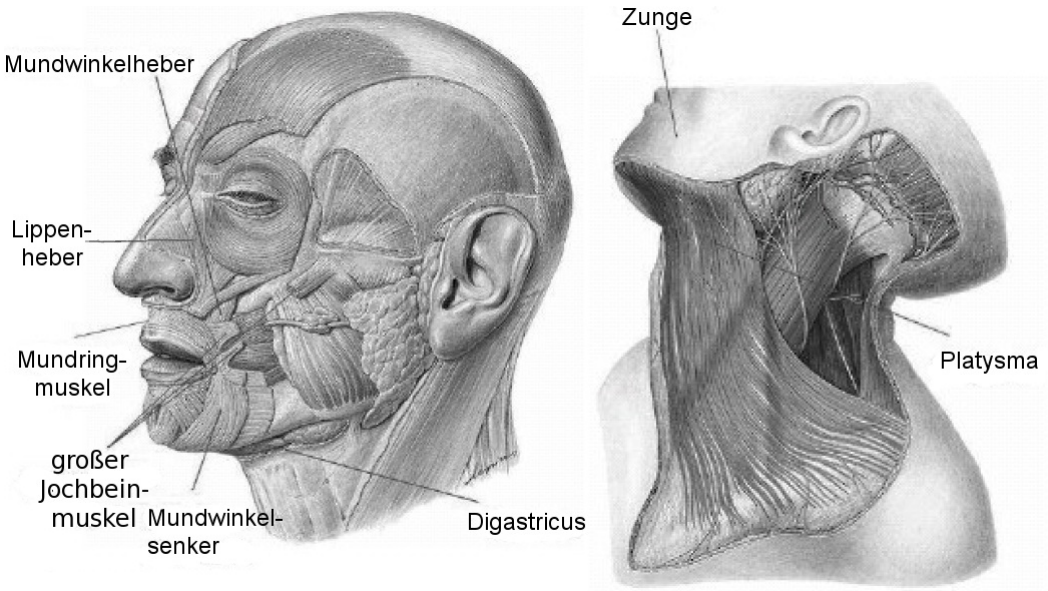


Abbildung 2.2: An der Spracherzeugung beteiligte Gesichtsmuskeln

Ein weit verbreiteter Ansatz ist es also die Signale der wichtigsten Muskeln mit einzelnen Elektroden zu erfassen. Lena Maier-Hain [8] hat dazu im Jahr 2005 ausführliche Studien betrieben. Einen anderen Ansatz wählt Christopher Schulte [11], der in seiner Arbeit von 2011 erstmals Elektrodenarrays in der EMG-basierten Spracherkennung vorstellt. Er verwendet Arrays verschiedener Größen, von denen er je eins an einer Wange und eins unter dem Kinn eines Probanden anbringt. Dabei erzielen Systeme mit Elektrodenarrays vergleichbare Ergebnisse zu Vorgängersystemen.

## 2.3 Das EMG-basierte Spracherkennungssystem

Basierend auf diesen Erkenntnissen kommen in dem für diese Arbeit verwendeten EMG-basierten Spracherkennungssystem Elektrodenarrays 3.3 zum Einsatz. Der Aufbau dieses Systems ist in Abbildung 2.3 skizziert.

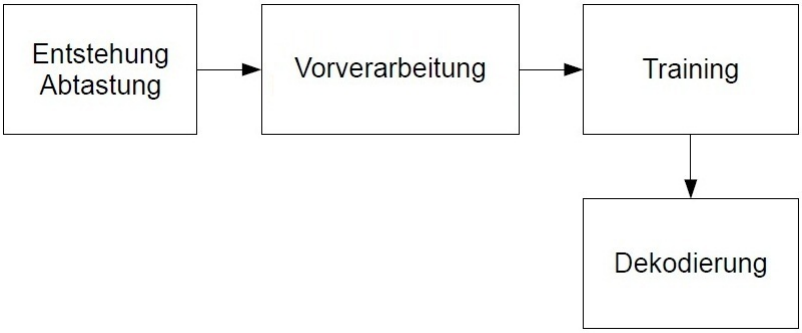


Abbildung 2.3: Abbildung: Schema des EMG-basierten Spracherkennungnssystems [11]

Zur Entstehung des EMG-Signals wurde bereits in den vorhergehenden Abschnitten die Muskelkontraktion und die dabei entstehenden Potentiale beschrieben. Zur Aufnahme des Signals werden dem Probanten Elektrodenarrays 3.3 im Gesicht angebracht, die mit einem Verstärker verbunden sind.

Aus den so gewonnenen Messdaten werden Merkmale durch die Methode der Time Domain Features [12] extrahiert, d.h. Merkmale werden abhängig vom normalisierten Mittel und Energie des Signals sowie der Zero-Crossing-Rate berechnet. Für jeden der fünf Elektrodenkanäle wird auf diese Weise ein Time Domain Feature berechnet. Den finalen Featurevektor erhält man schließlich durch Stacking.

Um die Dimension der Merkmale zu reduzieren, wird die Lineare Diskriminanzanalyse (LDA), eine Methode zur Reduktion der Dimensionalität der extrahierten Merkmale, sowie die Hauptkomponentenanalyse (PCA), eingesetzt. Die in dieser Arbeit entwickelten Interpolationsmethoden wurden nach der Berechnung der Time Domain Features eingesetzt.

Das Erkennungssystem verfolgt in der Modellierung zwei Ansätze: Den phonembasierten Ansatz und einen Ansatz basierend auf Bundled Phonetic Features (BDPF). Ein Phonem ist die kleinste Spracheinheit, die einen semantischen Unterschied zwischen zwei Wörtern bedeutet. Dem entsprechend ist es das Ziel des phonembasierten Ansatzes, Wahrscheinlichkeiten zu ermitteln mit denen Merkmale Phonemen zugeordnet werden können. Oft wird auch der engere Kontext mit betrachtet, um bessere Aussagen treffen zu können. Die BDPF-basierte Variante verwendet zusätzlich Phonetic Features, d. h. phonetische Informationen über die Konfiguration der Artikulatoren wie z.B. stimmhaft oder frikativ. Abhängige Features können zu Bündeln zusammengefasst werden, wie runder vorderer Vokal.

Der Erkennungsschritt wird mit Hidden Markov Modells (HMMs) bzw. dem Viterbi-Algorithmus umgesetzt. HMMs sind zweifach stochastische Modelle, bei denen sowohl die Zustandsfolge als auch die Emission einer Beobachtung in einem Zustand probabilistisch ist. Der Viterbi-Algorithmus ermittelt basierend darauf die wahrscheinlichste Folge.

# 3. Durchführung des Experiments

Dieses Kapitel geht auf die einzelnen Schritte in der Messkette des EMG-basierten Erkennersystems ein. Nach einem Blick auf den verwendeten Datenkorpus, sowie die bei der Messung eingesetzte Aufnahmesoft- und -hardware, wird der Fokus auf der Vorverarbeitung bzw. den zwei Erkennervarianten in BioKIT und Janus liegen.

## 3.1 Verwendeter Korpus

Für die Experimente mit Signalinterpolation wurden sieben Aufnahmesessions von sechs Sprechern verwendet. Dabei kamen 40 Elektroden in zwei Arrays - 32 Elektroden in einem Array an der Wange und acht Elektroden in einem Array unter dem Kinn - zum Einsatz. In Tabelle 3.1 sehen Sie eine Übersicht über die Wortfehlerraten der einzelnen Sessions, die ohne Interpolation ermittelt wurden. Diese dienen als Vergleichswerte für die im Laufe dieser Arbeit entwickelten Methoden.

| SpeakerID | SessionID | Wortfehlerrate (mit PCA) |
|---|---|---|
| 551 | 171 | 32,3% |
| 601 | 063 | 26,3% |
| 601 | 064 | 29,3% |
| 703 | 001 | 74,7% |
| 704 | 002 | 59,6% |
| 705 | 001 | 36,4% |
| 706 | 001 | 56,6% |

Tabelle 3.1: Wortfehlerraten der einzelnen Sprecher/Sessions ohne Interpolation

Die verwendeten Satzlisten wurden aus dem EMG-PIT-Korpus übernommen und wurden auch im EMG-UKA-Korpus eingesetzt. Die folgenden zehn Sätze wurden zum Testen verwendet:

- YOU MUST REMEMBER THAT CONSTRUCTIVE CRITICISM WILL MAKE US STRONGER

- THE OUTAGES WERE APPARENTLY CAUSED BY SYSTEM FAILURE NOT SABOTAGE
- AGENCY MANAGERS SAY TEMPORARY WORKERS ARE VALUABLE
- THE SUCCEEDED IN DOING THAT WITH A VENGEANCE
- THEY EVEN KNEW THAT THEY WERE RUNNING INTO THE MOUNTAIN
- THE STATE OF FLORIDA HAS A TOUGH POLICY AGAINST AMBULANCE CHASING
- THE CONVICTED MURDERER HAS AVOIDED EXECUTION BY LODGING REPEATED APPEALS
- MANY STATES DEVISE SOLUTIONS TO PROBLEMS OF WELFARE AND HEALTH CARE
- THE PENTAGON ADMITS INTERNET ATTACKERS ARE AN INCREASING THREAT TO COMPUTERS
- WHAT FOLLOWS IS AN OLD STORY MORE THAN FORTY YEARS OLD

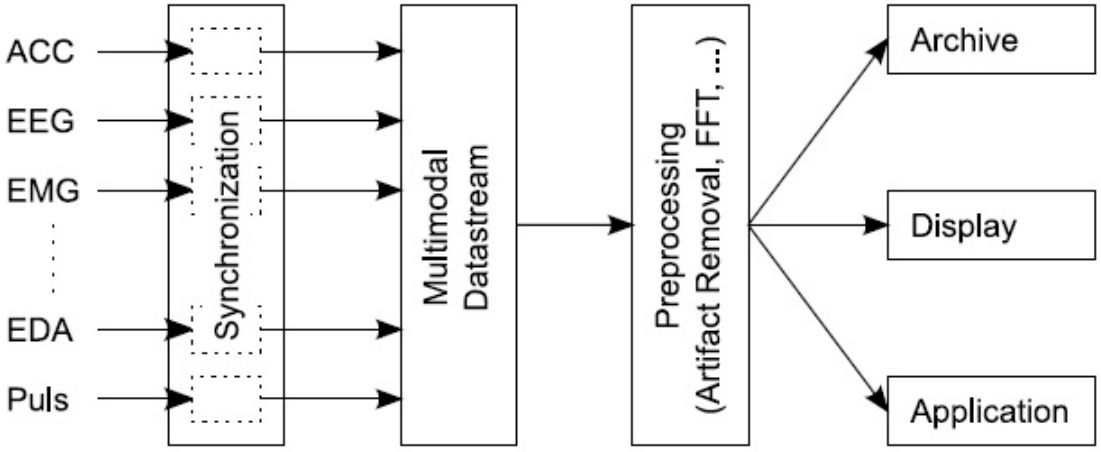


Abbildung 3.1: Modulare Architektur von Biosignals Studio aus [3]

## 3.2 Beschreibung der Aufnahmesoftware

Zur Aufnahme dieses Korpus kam Biosignals Studio (BSS, [3]) zum Einsatz. BSS ist ein Framework zur Sammlung multimodaler Sensordaten. Als Teil eines Mensch-Maschine-Interaktionssystems (human-machine interaction system HMI) ist BSS sowohl in der Entwicklungsphase nützlich, um große Mengen von Daten zu erfassen, als auch in der Anwendungsphase mit Echtzeitanforderungen. Dabei ist die Architektur

3.1 von BSS so entworfen, dass BSS nicht nur Signale von verschiedenen Quellen wie EMG, EEG, ACC usw. verarbeiten kann, sondern auch das Erstellen von neuen Applikationen möglichst einfach macht.

Dazu verwendet BSS Module, die jeweils für eine spezifische Aufgabe entworfen sind. Diese feingranulare Aufgabentrennung erlaubt ein großes Maß an Flexibilität, so dass Module einfach ausgetauscht werden und z.B. Ausgabeformate gewechselt oder verschiedene Algorithmen angewandt werden können, ohne das ganze System ändern zu müssen. Eine neue Anwendung kann dann einfach durch das Verknüpfen der Datenströme der jeweiligen Module erstellt werden.

## 3.3 Beschreibung der Hardware

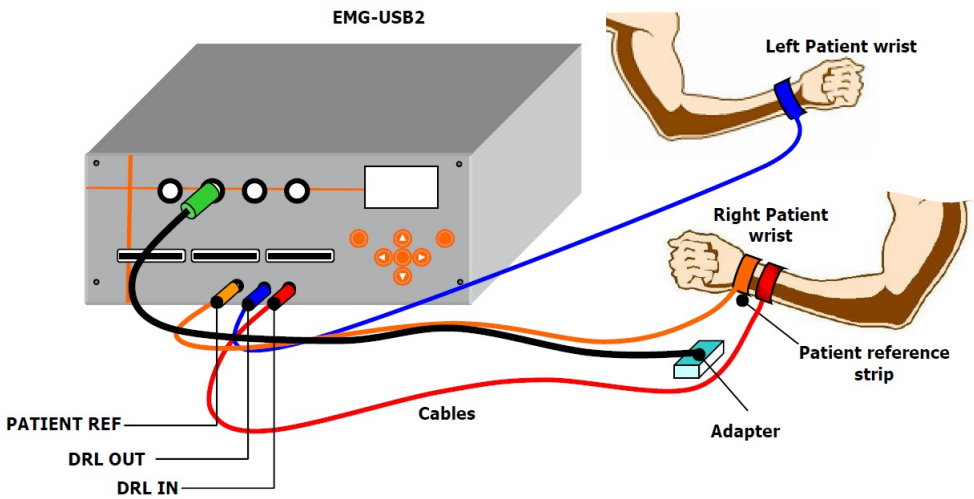


Abbildung 3.2: Darstellung des Versuchsaufbaus [10]

Als nächstes soll in diesem Abschnitt der Versuchsaufbau und die Hardware zum Sammeln von Messdaten beschrieben werden. Kernstück der Aufnahmetechnik ist der EMG-Verstärker EMG-USB2 der Firma OTBioelektrika, der intramuskuläre und extramuskuläre EMG-Signale empfangen und verarbeiten kann. Eine genaue Beschreibung des Verstärkers ist in der Arbeit von Christoph Schulte [11] oder auf der Website von OTBioelektrika [9] zu finden.

Wie in Abbildung 3.3 dargestellt wurde die Versuchsperson durch drei befeuchtete Stoffarmbänder mit dem Verstärker verbunden, um Störungen zu reduzieren bzw. um die Versuchsperson zu erden. Die Armbänder wurden so angebracht, dass sie sich nicht gegenseitig berührten und sich an Stellen befinden, die während der Aufnahme keine Muskelaktivität zeigten. Außerdem verwendete die Versuchsperson ein Headsetmikrofon, das parallel Audiodaten aufzeichnete.

Schließlich wurden die Elektrodenarrays über einen Vorverstärker (Adapter) mit dem EMG-USB2 verbunden. So konnten die Arrays dann mit adhäsiven Schaumstoffpads beklebt werden, die einen Einfüllbereich für die Elektrolytpaste und besseren Halt

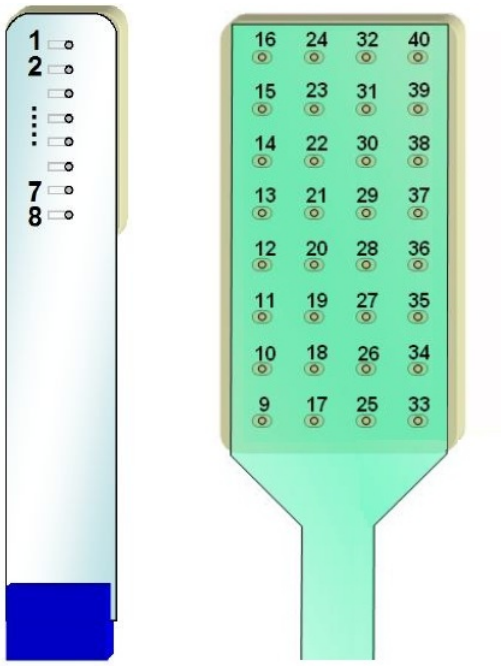


Abbildung 3.3: Schematische Darstellung der Elektrodenarrays. Verwendet wurden ein 8-Elektrodenarray (links) unter dem Kinn und ein 32-Elektrodenarray (rechts) an Wange des Probanten.

an der Haut bieten. Zuletzt wurden die Arrays im Gesicht 3.4 der Versuchsperson angebracht. Von den zwei Arrays wurde das kleinere mit acht Elektroden in einer Reihe unter dem Kinn, das zweite größere mit 32 Elektroden in vier Reihen an der Wange der Versuchsperson befestigt.

Abbildung 3.4: Abbildung der Elektrodenarrays an Probant.

## 3.4 Vorverarbeitung in mit/ohne PCA

Basierend auf den gesammelten Rohdaten wurden die folgenden drei Schritte als Vorverarbeitung durchgeführt:
1. Berechnung der Time Domain Features
2. Durchführen des LDA-Algorithmus
3. Durchführen des PCA-Algorithmus

### 3.4.1 Berechnung der Time Domain Features

Sind die EMG-Signale gemessen, müssen aus den Rohdaten die Merkmale extrahiert werden. Die hierzu verwendete Methode ist die Berechnung von "Time Domain Features"[12]. Diese beinhaltet folgendes Vorgehen: Berechne für jedes Mermal dessen Frame-basiertes Mittel im Zeitraum, dessen Frame-basierte Energie und dessen Frame-basierte Zero-Crossing-Rate. Außerdem wird aus dem EMG-Signal ein begradiktes Hochfrequenz-Signal berechnet, so dass das Time Domain Feature als Stapel adjazenter Frames abhängig von den genannten Parametern definiert werden kann.

### 3.4.2 Durchführung des LDA-Algorithmus

Bis jetzt besteht allerdings noch das Problem, das sowohl Rohdaten als auch interpolierte Daten eine sehr hohe Dimensionalität aufweisen. Um diese auf nur noch 32 Dimensionen zu reduzieren, wird eine Lineare Diskriminanzanalyse (LDA) verwendet. Das Ziel dieser in der Spracherkennung weit verbreiteten Methode zur Merkmalsreduktion ist es, Merkmale, die Störungen oder nutzlose Informationen enthalten, heraus zu filtern, ohne dabei Informationen zu verlieren, die zur Merkmalsunterscheidung nützlich sind.

Eine Einschränkung ist, dass die Klassenzugehörigkeit der Daten schon vor Anwendung der LDA bekannt sein muss. Für n zu unterscheidende Klassen werden also auch n Datensätze benötigt, die jeweils die Daten einer Klasse enthalten.

Um die Dimensionalität zu reduzieren berechnet der LDA-Algorithmus dann eine Transformationsmatrix, die auf die Eingabedatensätze angewandt, zu neuen nach Diskriminanz geordneten Dimensionen führt. Hiervon wählen wir die ersten transformierten Dimensionen aus und erreichen damit unser Ziel: Reduzierte Dimensionalität und weniger Störeinflüsse.

### 3.4.3 Durchführung des PCA-Algorithmus

Um die Erkennerleistung weiter zu verbessern kann zusätzlich zum LDA-Algorithmus auch die Hauptkomponentenanalyse (Principal Component Analysis, PCA) verwendet werden. Im Gegensatz zum LDA-Algorithmus sucht die PCA sukzessiv zueinander orthogonale Richtungen mit großer Varianz und liefert diese geordnet nach Varianz zurück. Dabei wird angenommen, dass Richtungen mit größter Varianz auch von großem Interesse sind und Richtungen mit geringer Varianz eventuell nur Störungen bzw. Rauschen enthalten. Richtungen mit großer Varianz werden auch Hauptkomponenten genannt. Sind diese gefunden, wird im Featureraum ein Basiswechsel durchgeführt, so dass die neue Basis aus Hauptkomponenten besteht. In dieser Basis dargestellt, sind die Komponenten der transformierten Vektoren nach Varianz geordnet. Will man Kompression durchführen, können dann einfach die letzten Komponenten mit geringster Varianz weggelassen werden.

In dieser Arbeit werden die Auswirkungen von Signalinterpolation sowohl mit als auch ohne PCA in der Vorverarbeitung untersucht.

## 3.5 Erkenner

Als letzter Schritt der Vorverarbeitung werden die extrahierten Merkmale normalisiert, so dass je eine Transformationsmatrix aus dem LDA-Algorithmus und dem PCA-Algorithmus vorliegen und zusammen mit den extrahierten Merkmalen an einen Erkenner übergeben werden können.

### 3.5.1 Janus

Das Janus Recognition Toolkit (JRTk) ist eine universelle Werkzeugsammlung zur Spracherkennung. Mit Hilfe des JRTk wurde hier am CSL ein Spracherkenner entwickelt, der im Folgenden der Einfachheit halber Janus-Erkenner genannt wird. Der Janus-Erkenner bekommt als Eingabe eine LDA-Matrix, zwei Sätze vorverarbeitete Daten und je nach Verwendung auch eine PCA-Matrix. Von den zwei Datensätzen ist der erste ganz ohne Signalinterpolation vorverarbeitet, während im zweiten jeder Kanal interpoliert ist. Dies ermöglicht es das Training nur einmal durchzuführen und dann für n Kanäle die Testphase n-mal durchzuführen, um dabei je einen Kanal aus dem ursprünglichen Datensatz durch einen interpolierten Kanal zu ersetzen. Auf diese Weise können effizient die Auswirkungen der Interpolation einzelner Kanäle auf die Erkennerleistung untersucht werden.

Für hörbar gesprochene Sprache ist es nötig, vor Beginn des Trainings Time-Alignments zu berechnen, d.h. es wird eine Zuordnung zwischen Signalfenstern und Phonemklassen auf Basis der akustisch parallel aufgenommenen Daten hergestellt. Der Erkennungsschritt teilt sich dann unter Verwendung von Bundled Phonetic Features (BDPFs) in die folgenden drei Schritte:

In der ersten Trainingsphase wird eine LDA-Matrix auf den Eingabedaten berechnet. Als Modelle werden zu Beginn kontextunabhängige Phoneme sowie binärwertige Phonetic Features verwendet. Als nächstes werden mit einem Merge-and-Split-Algorithmus initiale Modelle berechnet, auf denen dann vier Iterationen Viterbi-Training ausgeführt wird.

Unter Verwendung dieser Modelle kann in der zweiten Phase das Phonetic Feature Bundling durchgeführt werden.

Wie in der ersten Phase wird in der dritten und letzten Trainingsphase eine LDA-Matrix berechnet. Es wird ein Merge-and-Split-Algorithmus zum Berechnen für initiale Modelle verwendet und darauf wieder vier Iterationen Viterbi-Training durchgeführt. Im Unterschied zur ersten Phase werden die Phonemmodelle nicht mehr beachtet. Stattdessen werden jetzt nur noch die BDPF-Modelle verwendet.

In der Testphase werden, wie oben erwähnt, einzelne Kanäle aus dem ursprünglichen Datensatz durch interpolierte Kanäle ersetzt, so dass sich für n Kanäle n Datensätze mit je einem interpolierten Kanal ergeben. Auf diesen Datensätzen wird dann das Testen durchgeführt.

Für den Janus-Erkenner wurde sowohl Vorverarbeitung mit PCA also auch ohne untersucht.

### 3.5.2 BioKIT

Im Gegensatz zum Janus-Erkenner, mit dem schon seit einigen Jahren am CSL Forschung betrieben wird, kommt das Biosignal Recognition Toolkit (BioKIT) erst seit kurzem in der Spracherkennung zum Einsatz. BioKIT ist ein vielseitig einsetzbares System zur Verarbeitung und Dekodierung von Biosignalen. Es enthält eine Vielzahl von Vorverarbeitungsoperatoren, bietet effiziente HMM-basierte Sequenzmodellierung und -dekodierung sowie ein modulares state-of-the-art Design mit einem Python basierten Skriptinterface.

Da in BioKIT noch keine Bundled Phonetic Features implementiert sind, beschränkt sich der Erkenner hier auf ein Merge-and-Split-Training zur Erstellung initialer Modelle, ein Viterbi-Training, das darauf ausgeführt wird und schließlich ein Schritt zur Dekodierung.

# 4. Experimente und Resultate

Dieses Kapitel enthält die Vorstellung eines naiven Lösungsansatzes zur Signalinterpolation, der schließlich durch einen Maximum-Likelihood-Algorithmus erweitert wird. Das Ziel dieses Algorithmus ist es, Signalinterpolation gewichtet durchzuführen, so dass die Unterschiede zwischen interpoliertem und ursprünglichen Signal minimiert werden. Weiterhin stellt das Kapitel Experimente vor, auf denen der Algorithmus durchgeführt wurde. Dabei kamen Ekenner in BioKIT und Janus unter verschiedenen Konfigurationen zum Einsatz, um ein Maß für die Güte des Algorithmus zu erhalten. Abschließend folgt eine Analyse der so gewonnen Ergebnisse.

## 4.1 Signalinterpolation

Um die Frage nach den Auswirkungen der Signalinterpolation beantworten zu können, soll diese hier erst genau untersucht werden. Bei Signalinterpolation geht es allgemein darum einen Kanal innerhalb eines Elektrodenarrays durch Werte zu ersetzen, die aus den Signalen der Kanäle ermittelt werden, die sich in der Nachbarschaft des zu ersetzenden Kanals befinden. Dabei ist klar, dass der Rand des Arrays eine Einschränkung für die Nachbarschaft bedeutet. Figur 4.1 stellt eine vier-Kanal-Nachbarschaft innerhalb und am Rande eines Elektrodenarrays dar.

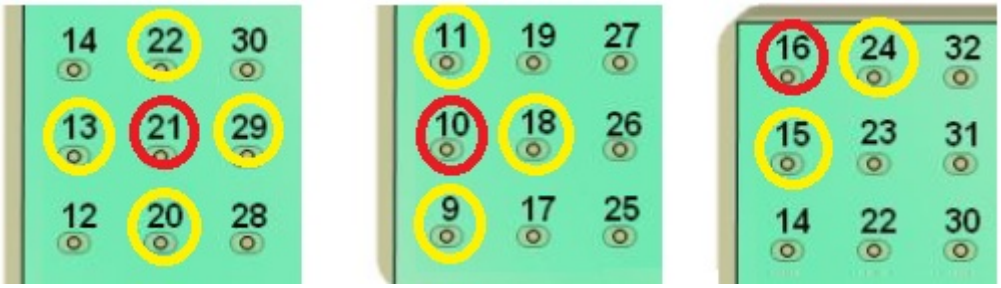


Abbildung 4.1: Veranschaulichung einer 4-Kanal-Nachbarschaft innerhalb (links), am Rand (mitte) und in einer Ecke (rechts) eines Elektrodenarrays

Auf den Trainingsdaten soll nun ein interpoliertes Signal aus einer Nachbarschaft berechnet werden. Dazu war der erste Schritt eine naive Berechnung eines Mittelwerts der Signale aus der Kanalnachbarschaft durchzuführen. Die Formel zu dieser "naiven Interpolation" sieht wie folgt aus:

Seien $n_1$(t), ..., $n_N$(t) die Signale der Nachbarelektroden des zu interpolierenden Signals abhängig von der Zeit t und N deren Anzahl. Dann gilt für das interpolierte Signal p(t)

$$p(t) = \frac{1}{N} \sum_{i=1}^{N} n_i(t) \tag{4.1}$$

Der nächste Schritt war es, diese Methode zu verfeinern. Da nicht davon auszugehen ist, dass eine gleichmäßige Beteiligung aller Kanäle aus der Nachbarschaft an der Signalinterpolation immer ideal ist, führen wir als nächstes die "gewichtete Interpolation" ein. Durch das Gewichten von Signalen bestimmter Kanäle, soll es möglich werden, dass manche Kanäle in der Nachbarschaft stärker berücksichtigt werden, um so ein besseres Ergebnis zu erzielen. "Besser" heißt in diesem Fall, dass das berechnete Signal näher am ursprünglichen Signal ist. Denkbar wäre dies z.B. wenn ein Kanal, der die selbe Muskelfaser erfasst wie der zu ersetzende Kanal, stärker gewichtet wird als ein Kanal auf einer anderen Muskelfaser, oder wenn ein Kanal, der weniger Artefakte aufweist, bevorzugt wird.

Dazu sei s(t) das Signal des zu ersetzenden Kanals und $n_1$(t), ..., $n_N$(t) wieder die Signale der Nachbarkanäle. Außerdem seien $w_1$, ..., $w_N$ die zu berechnenden Gewichte. Dann soll für das interpolierte Signal p abhängig vom Zeitpunkt t gelten

$$p(t) = \sum_{i=1}^{N} w_i \cdot n_i(t) \tag{4.2}$$

Ziel ist es dann den quadratischen Fehler zu minimieren, der entsteht, wenn s(t) durch p(t) ersetzt wird. Der Ansatz sieht also wie folgt aus:

$$T(w_1, \ldots, w_N) = \frac{1}{N} \sum_{t} (s(t) - p(t))^2 \tag{4.3}$$

Dabei ist der quadratische Fehler T gleich der mittleren Abweichung des interpolierten Signals vom Originalsignal abhängig von den Gewichten $w_1$, ..., $w_N$. Diese gilt es so zu bestimmen, so dass T minimimal wird. Wir ersetzen also p(t) nach Gleichung (4.2) und differenzieren:

$$\frac{d}{d\vec{w}} T(w_1, \ldots, w_N) = \frac{d}{d\vec{w}} \sum_{t} (s(t) - \sum_{i=1}^{N} w_i \cdot n_i(t))^2 \tag{4.4}$$

$$
= \begin{pmatrix} \sum_t [2(s(t) - \sum_{i=1}^{N} w_i \cdot n_i(t)) \cdot (n_0(t)] \\ \sum_t [2(s(t) - \sum_{i=1}^{N} w_i \cdot n_i(t)) \cdot (n_1(t)] \\ \dots \\ \sum_t [2(s(t) - \sum_{i=1}^{N} w_i \cdot n_i(t)) \cdot (n_N(t)] \end{pmatrix}
$$

$$
= \sum_t 2(s(t) - \sum_{i=1}^{N} w_i \cdot n_i(t)) \cdot \begin{pmatrix} -n_0(t) \\ -n_1(t) \\ \dots \\ -n_N(t) \end{pmatrix}
$$

Diese Formel setzen wir gleich null und lösen das resultierende Gleichungssystem, um $w_1, \dots, w_N$ zu erhalten. Mit den so bestimmten Gewichten ist der Abstand des interpolierten Signal zum Originalsignal minimal. Da das interpolierte Signal aus den Nachbarsignalen berechnet wird, werden Störungen, die im Originalsignal aber nicht in den Nachbarsignalen auftreten, behoben, ohne das Signal zu stark zu verändern. Damit erhalten wir eine Methode, um Kanäle zu interpolieren, die an den Anfang der Vorverarbeitung gesetzt und direkt auf den Rohdaten ausgeführt werden kann.

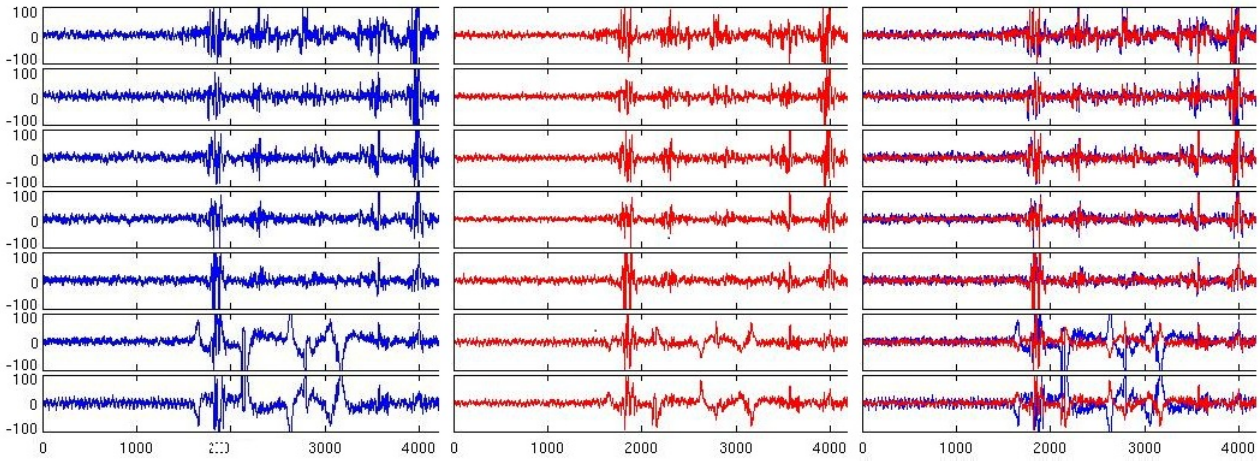


Abbildung 4.2: Beispiel für EMG-Signale (blau) und die dazu gehörigen (gewichtete) interpolierten Signale (rot) von Sprecher 601, Aufnahme 063, Utterance 15, Kanäle 17 bis 23 (Zählung wie in Abbildung 3.3)

Abbildung 4.2 zeigt beispielhaft die Signale von Kanäle aus Aufnahme 063, sowie die Signale nach der Durchführung der gewichteter Interpolation. Deutlich sichtbar ist, dass die Artefakte in Kanal 22 und 23 abgeschwächt werden.

Um herauszufinden, ob die Vorzüge der gewichteten Interpolation gegenüber der Naiven den höheren Rechenaufwand rechtfertigen, werden beide Methoden in den nächsten Abschnitten ausführlich bewertet.

## 4.2 Auswertung durch Spracherkennung

Um die an die Signalinterpolation gestellten Erwartungen zu evaluieren, wurde sie auf dem Korpus 3.1 ausgeführt. In den nächsten zwei Abschnitten sollen die daraus resultierenden Ergebnisse vorgestellt werden. Untersucht wurden naive und gewichtete Interpolation mit Nachbarschaften von vier bzw. acht Kanälen. Als Maßstab für die Güte der Interpolation wird die Wortfehlerrate (Word-Error-Rate WER), ermittelt durch die Erkenner in Janus und BioKIT, verwendet. Die Signalinterpolation wurde unter einfachen Bedingungen im phonembasierten System, das in BioKIT implementiert ist, ausgewertet und auch auf dem umfangreicheren System in Janus, das zusätzlich Bundled Phonetic Features implementiert hat, evaluiert. Die Vorverarbeitung wurde sowohl mit als auch ohne PCA durchgeführt, um auch dort Auswirkungen der Signalinterpolation feststellen zu können.

### 4.2.1 Phonembasierte Auswertung mit BioKIT

Zu Beginn der Versuchsreihe mit BioKIT wurden der Erkenner auf den Eingabedaten ohne Interpolation laufen gelassen. In der Vorverarbeitung wurde der PCA-Algorithmus verwendet. Um das Verhalten der Interpolationen ohne PCA zu untersuchen, sollte noch abgewartet werden, welches der beiden Systeme die besseren Werte liefert und schließlich fiel die Entscheidung für weitere Untersuchungen auf den Janus-Erkenner. Die Ergebnisse aus den Berechnungen mit dem BioKIT-Erkenner ohne Interpolation sehen dann wie folgt aus:

| SpeakerID | SessionID | Wortfehlerrate (mit PCA) |
|---|---|---|
| 551 | 171 | 69,70% |
| 601 | 063 | 64,64% |
| 601 | 064 | 58,58% |
| 703 | 001 | 84,84% |
| 704 | 002 | 77,78% |
| 705 | 001 | 71,71% |
| 706 | 001 | 82,82% |

Abbildung 4.3: Wortfehlerraten der einzelnen Sessions ohne Interpolation für den BioKIT-Erkenner mit PCA.

Tabelle T.1 zeigt die Ergebnisse der Interpolationsberechnungen unter Verwendung des BioKIT-Erkenners und ist wie folgt aufgebaut: Auf der horizontalen Achse ist die jeweilige Durchführung aufgetragen, gekennzeichnet durch die Interpolationsmethode (Naiv mit 4- oder 8-Elektrodennachbarchaft und Gewichtet mit 4- oder 8-Elektrodennach- barchaft) sowie die SpeakerID und die SessionID. Dagegen ist auf der vertikalen Achse der jeweils interpolierte Kanal eingetragen, so dass man schließlich die resultierende WER ablesen kann. Die Mittelwerte und Minima sind in Tabelle A.1 aufgetragen.

### 4.2.2 BDPF-basierte Auswertung mit Janus

Die Vorgehensweise für den Janus-Erkenner war zu der mit dem BioKIT-Erkenner sehr ähnlich. Zunächst wurde wieder auf Interpolation verzichtet, um so eine Vergleichsbasis zu schaffen. Auf diese Weise erhielten wir die in Tabelle 4.4 dargestellten

WER. Zusätzlich sind nicht nur WER mit, sondern auch ohne PCA in der Vorverarbeitung abgebildet. Es fällt gleich auf, dass die WER mit PCA deutlich niedriger sind als ohne.

| SpeakerID | SessionID | Wortfehlerrate (onhe PCA) | Wortfehlerrate (mit PCA) |
|---|---|---|---|
| 551 | 171 | 41,4% | 32,3% |
| 601 | 063 | 47,5% | 26,3% |
| 601 | 064 | 42,4% | 29,3% |
| 703 | 001 | 92,9% | 74,7% |
| 704 | 002 | 62,6% | 59,6% |

Abbildung 4.4: Wortfehlerraten der einzelnen Sessions ohne und mit PCA in der Vorverarbeitung

Außerdem sind die Sprecher 705 und 706 noch gesondert zu erwähnen. In den Aufnahmensessions dieser beiden Sprecher fanden sich Störungen, die dazu führten, dass im Viterbibeamsearch-Algorithmus der Beam überschritten wurde und somit keine finale Hypothese bestimmt werden konnte. Ein systematisches Testen war damit nicht möglich. Stattdessen sollen Sprecher 705 und 706 als Beispiel dienen, wie unbrauchbare Sessions durch Signalinterpolation wieder nutzbar gemacht werden konnten. Dazu wurden die Sessions manuell auf Störungen untersucht, um die entsprechenden Kanäle dann interpolieren zu lassen. Dieses Phänomen trat allerdings nur im BDPF-basierten System auf.

Auf den übrigen Sessions konnte der Januserkenner durchgeführt werden. Dabei wurde jeweils ein Kanal interpoliert und ausgetauscht, so dass das Verhalten der Interpolation für einzelne Kanäle deutlich wird. Tabelle T.2 im Anhang zeigt die daraus resultierende Ergebnisse ohne PCA in der Vorverarbeitung, während Tabelle T.3 Ergebnisse aus Berechnungen mit PCA in der Vorverarbeitung darstellt.

Beide Tabellen sind aufgebaut wie die entsprechende Tabelle T.1 im vorherigen Abschnitt.

Die aus diesen Berechnungen resultierenden Mittelwerte und Minima, also die niedrigsten Wortfehlerraten, zeigt Tabelle A.2.

### 4.2.3 Vergleich Janus/BioKIT

Da der Janus-Erkenner für die Sprecher 705 und 706 nicht ohne Weiteres Ergebnisse lieferte, basiert dieser Vergleich auf den Sessions der übrigen Sprecher unter Verwendung der PCA.

Auf den ersten Blick ist dabei zu erkennen, dass der Janus-Erkenner deutlich bessere Mittelwerte liefert. Aufnahme 064 von Sprecher 601 (4.5) soll hier als unterstützendes Beispiel dienen. Als Vergleich können die Diagramme der übrigen Aufnahmen im Anhang (A.3) betrachtet werden. Der Grund für die niedrigeren Mittelwerte des

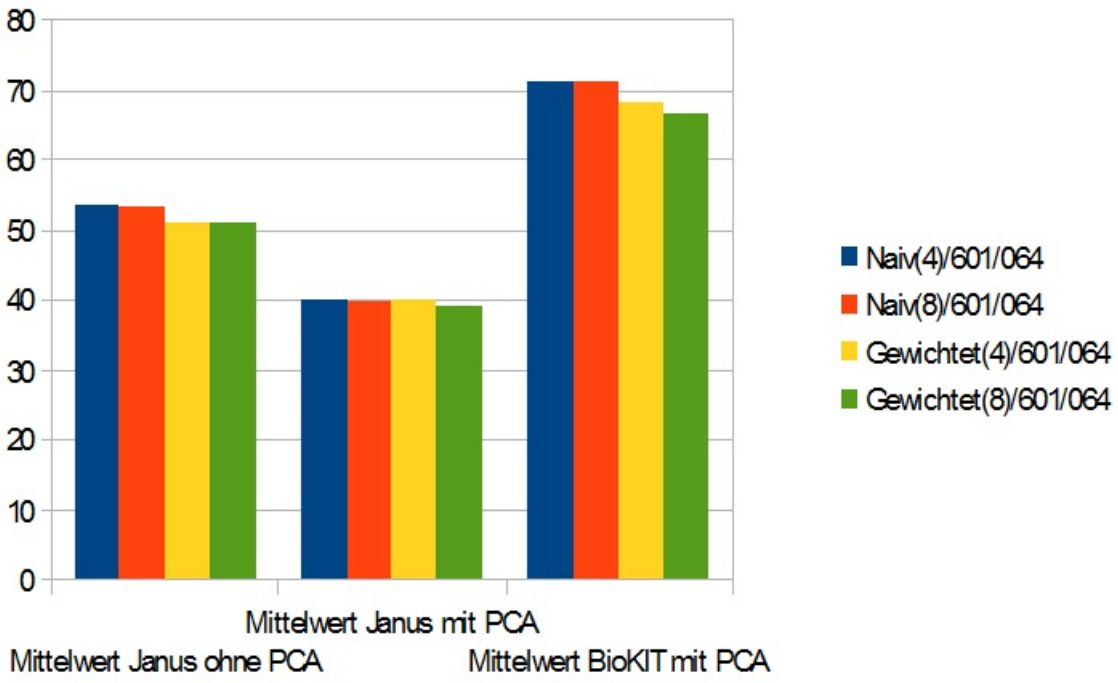


Abbildung 4.5: Mittelwerte von Sprecher 601, Aufnahme 064

Janus-Erkenners findet sich in der Verwendung der Bundled Phonetic Features ([12]).

Auffällig ist außerdem das breite Spektrum an Wortfehlerraten, die der Janus-Erkenner liefert. Neben Zahlen, die sogar unterhalb des Referenzwertes liegen, liefert er auch völlig unbrauchbare Ergebnisse. Dagegen unterliegen die Ergebnisse des BioKIT-Erkenners keinen so großen Schwankungen und liegen in einem schmalen Rahmen um den Referenzwert verteilt. Auch hier finden sich Wortfehlerraten unterhalb des Referenzwertes.

## 4.3 Analyse

Zur Analyse der Signalinterpolation werden die oben genannten Ergebnisse herangezogen. Die gewichtete Interpolation soll gegenüber der naiven Interpolation als beste Methode vorgestellt werden.

### 4.3.1 Beste Methode

Dieses Ergebnis ergibt sich wie folgt: Jede Session wurde vom Janus-Erkenner mit PCA und ohne PCA und vom BioKIT-Erkenner für jede Interpolationsmethode jeweils 35 Mal verarbeitet. Dabei wurde je einer der 35 Kanäle interpoliert, sodass pro Session und pro Methode 35 Wortfehlerraten berechnet wurden. Die daraus errechneten Mittel- und Minimalwerte, die in Tabelle A.1 und A.2 im Anhang dargestellt sind, dienen als Vergleichskriterium für die Interpolationsmethoden. Für eine anschaulichere Betrachtung werden hier aber auch beispielhaft die Mittelwerte einiger

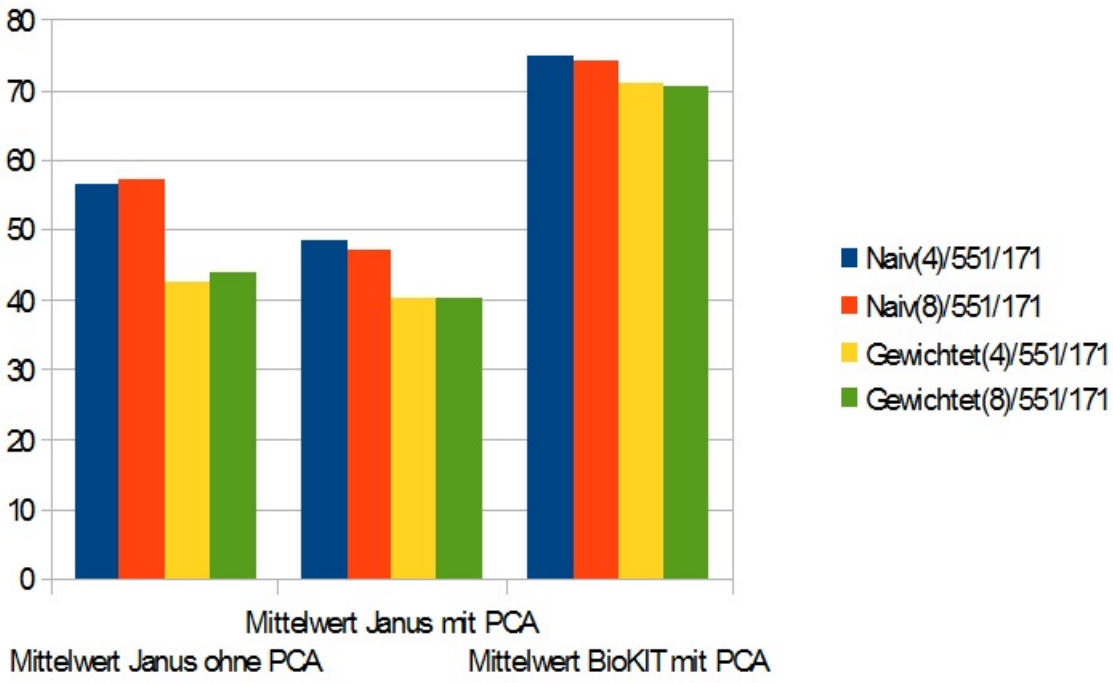


Abbildung 4.6: Mittelwerte von Sprecher 551, Aufnahme 171

Aufnahmen in Balkendiagrammen 4.6 dargestellt. Die Diagramme aller Aufnahmen befinden sich im Anhang A.3.

Deutlich wird vor allem, das die verschiedene Methoden ähnlich gute Ergebnisse erzielen. Bei näherer Betrachtung lassen sich jedoch folgende Beobachtungen machen:

Unter Berücksichtung der Mittelwerte ergibt sich ein, wenn auch oft nur knapper, Vorsprung für die gewichtete Interpolation. Für den Janus-Erkenner ohne PCA fallen die Unterschiede teilweise sogar deutlich aus. So erzielte die gewichtete Interpolation in der Session 704/002 4.7 z.B. eine relative Verbesserung von 5,6 % und in der Session 551/171 4.6 sogar eine Verbesserung von 24,3 % gegenüber der naiven Interpolation.

Der BioKIT-Erkenner schließt sich dieser letzten Einschätzung in abgeschwächter Form an. Durch die fehlenden BDPF resultieren hohe Wortfehlerraten und nur geringe Unterschiede zwischen den Methoden. Diese deuten aber weiterhin auf eine Überlegenheit der gewichteten Interpolation hin.

Dem entsprechend fällt auch das abschließende Urteil zu Gunsten der gewichteten Interpolation aus. Eine größere Nachbarschaft von acht Elektroden macht nochmal eine kleine Verbesserung von ca. 0.25% bei den Mittelwerten und liefert damit kein stichhaltiges Argument für die Verwendung der größeren Nachbarschaft. Anbetracht dieser verschwindend geringen Verbesserung schien es nicht sinnvoll noch

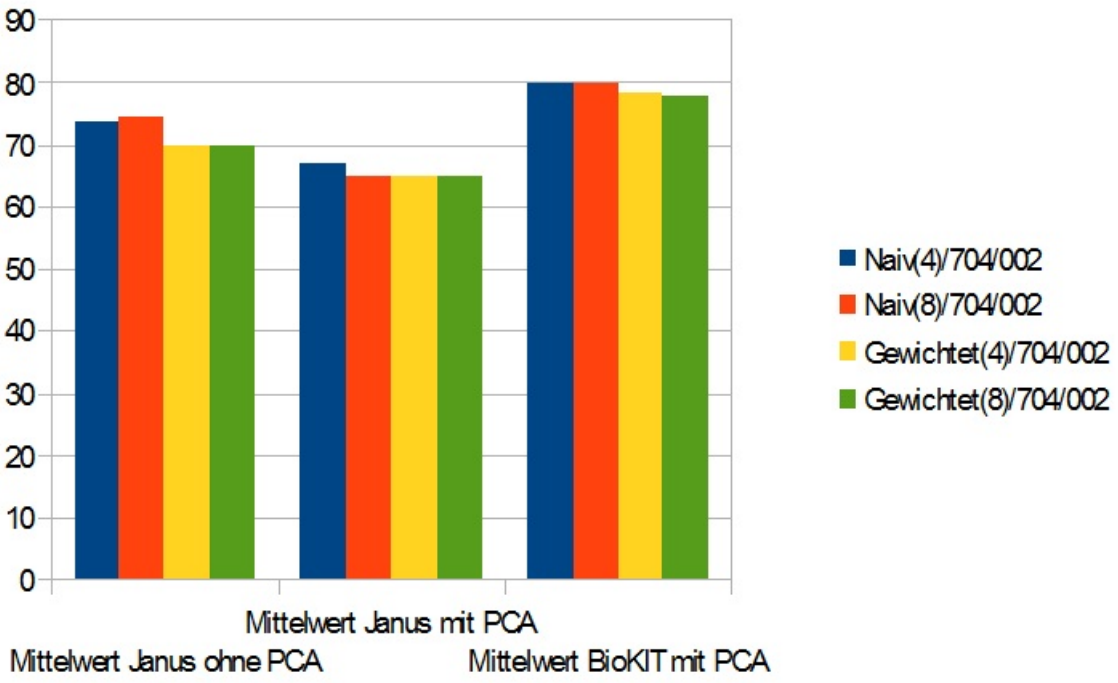


Abbildung 4.7: Mittelwerte von Sprecher 704, Aufnahme 002

größere Nachbarschaften von etwa 16 Elektroden zu untersuchen, da diese allein schon wegen der zu geringen Größe der Elektrodenarrays an keiner Position des Arrays vollständig möglich wären.

Zuletzt sei am Ende dieses Abschnitts noch erwähnt, dass zwar die Mittelwerte oft über den Referenzwerten liegen, die Minima, also die besten Werte aus den jeweiligen Aufnahmen, aber mindestens gleich den Referenzwerten sind und oft noch darunter liegen, d.h. durch geschickte Auswahl des interpolierten Kanals könnte eine relative Verbesserung im einstelligen Prozentbereich erzielt werden. Damit erfüllt die Signalinterpolation nicht nur die Anforderung, die Kanäle ohne große Verluste ersetzen zu können, sondern auch die Vermutung, die Ergebnisse noch verbessern zu können.

### 4.3.2 Besonderheiten

In diesem Abschnitt soll auf Besonderheiten eingegangen werden, die während des manuellen Untersuchens des Korpus aufgefallen sind. Ein Beispiel für gut funktionierende Interpolation findet sich in Aufnahme 601/063. In Kanal 22 und 23 dieser Aufnahme treten ungewöhnliche Ausschläge auf, die durch die Interpolation gedämpft werden. Die Folge ist eine Verbesserung der Wortfehlerrate.

Das Verhalten der Interpolation auf stark verrauschten Kanälen 4.8 ist allerdings nicht so eindeutig. Kanal 28 und 35 der Aufnahme 551/171 sind Beispiele für solche Kanäle. Während der Janus-Erkenner ohne PCA und der BioKIT-Erkenner nur geringe Verluste verzeichnen, sind diese beim Janus-Erkenner mit PCA deutlicher.

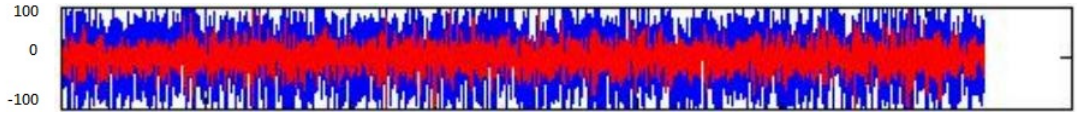


Abbildung 4.8: Beispiel für eine Störung im aufgezeichneten EMG-Signal von Sprecher 706, Aufnahme 001, Utterance 5, Kanal 3 (Zählung wie in Abbildung 3.3). Originalsignal (blau), Interpoliertes Signal (rot)

Die Vermutung, dass stark verrauschte Kanäle gut interpoliert werden können, kann also nur insofern bestätigt werden, als dass die Anforderung nach geringen Verlusten außer beim Janus-Erkenner mit PCA weitgehend erfüllt wird.

Abbildung 4.9: Beispiel für ein fehlendes EMG-Signal von Sprecher 704, Aufnahme 002, Utterance 5, Kanäle 2 (Zählung wie in Abbildung 3.3)

Es kann vorkommen, dass sich das Elektrodenarray vom Gesicht des Probanten löst und die entsprechenden Elektroden kein Signal bzw. ein Null-Signal 4.9 liefern. Die zugehörigen Kanäle sollten gut zu interpolieren sein. Dies trift nur teilweise zu. Während die naive Methode die Wortfehlerraten durch Interpolieren des achten Kanals in der Aufnahme 551/171 oder auch des achten oder neunten Kanals in der Aufnahme 704/002, die Wortfehlerrate deutlich erhöht, erzielt die gewichtete Variante Werte, die von der Referenz kaum abweichen. Daraus ist zu schließen, dass die gewichtete Interpolation die bessere Wahl ist, wenn sich z.B. Elektroden an den Ecken des Arrays vom Gesicht lösen und kein Signal mehr liefern.

Zuletzt sollen noch die Aufnahmen 705 und 706 betrachtet werden. Aufgrund starker Störungen schon ab dem neunten Kanal, war es nicht möglich eine finale Hypothese im Viterbibeamsearch-Algorithmus zu bestimmen. Zwar konnte eine Referenz berechnet werden, aber die Signalinterpolation konnte nicht systematisch getestet werden. Um trotzdem eine Durchführung zu ermöglichen, wurde der Ansatz gewählt, manuell störungsbehaftete Kanäle zu identifizieren und bereits in der Vorverarbeitung durch interpolierte Kanäle zu ersetzen. Das Ergebnis war für Aufnahme 705 eine Verbesserung der Wortfehlerrate im Mittel um 27,7% und für Aufnahme 26,9%. Auch wenn diese Zahlen sehr viel versprechend wirken, muss bedacht werden, dass es sich um nur zwei Aufnahmen handelt, die als Basis für ein aussagekrätiges Urteil nicht aussreichend sind. Sie sollten viel mehr als Anreiz gesehen werden, mehrfache Signalinterpolation weiter zu erforschen und mit einer automatischen Erkennung von störungsbehafteten Kanälen zu kombinieren. Da die Entwicklung eines solchen Automatismus den Rahmen dieser Arbeit übersteigt, können wir hier nur auf noch folgende Forschung verweisen.

### 4.3.3 Anwendungsszenario

Wir haben nun verschiedene Interpolationsmethoden kennen gelernt und ausgewertet. Die Frage mit der sich dieser letzte Abschnitt dieses Kapitels beschäftigt, ist die

Frage nach dem Anwendungsbereich der Interpolation.

Klar ist die Anwendung des EMG-basierten Spracherkenners. Etwa in Situationen wie in einer Bibliothek ist lautes Sprechen untersagt, in einem Zug will man nicht die anderen Fahrgäste stören oder in einem Geschäftsmeeting soll nicht jeder vertrauliche Informationen mithören. Menschen wiederum, denen durch körperliche Behinderung das Sprechen nicht mehr möglich ist, sind auf einen geeigneten Spracherkenner angewiesen, um trotzdem kommunikationsfähig zu bleiben.

Da sich all diese Anwendungen außerhalb eines abgeriegelten Labors befinden, ist ein solches Spracherkennungssystem einer Vielzahl von Störgrößen ausgesetzt, die von innerhalb oder außerhalb des Anwenders kommen. Diese können z.B. von Steckdosen oder offenen Anschlüssen an elektrischen Geräten stammen, aber auch von Muskelbewegungen, die unabhängig von der Spracherzeugung ablaufen, wie etwa Herzschlag oder Augenbewegung. Daraus resultiert die Forderung nach einem robusten Spracherkennungssystem, das nicht jedes Mal aussetzt, wenn der Anwender, in die Nähe einer Steckdose kommt oder mit den Augen blinzelt.

Ein Schritt in Richtung robustes Spracherkennungssystem ist die Signalinterpolation für Elektrodenarrays. Eingebettet in ein System, das gestörte Signale automatisch erkennt und anschließend interpoliert, könnte Signalinterpolation die Störanfälligkeit der EMG-basierten Spracherkennung senken. Wie weit dieser Ansatz schließlich im Alltag umsetzbar ist, bleibt Thema weiterer Forschung.

# 5. Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Verfahren zur Signalinterpolation für Elektrodenarrays innerhalb eines EMG-basierten Spracherkennungssystems entwickelt.

Es wurden mehrere Variante des Verfahrens entwickelt und auf sieben schon bestehenden Aufnahmesessions evaluiert. Dabei kamen aktuelle Spracherkenner zum Einsatz.

Es wurden Möglichkeiten aufgezeigt, wie durch Signalinterpolation Signale manipuliert werden können, ohne dabei die Erkennungsleistung zu verderben. Teilweise konnte diese sogar verbessert werden. Genauso wurde auch auf Einschränkungen der Interpolation hingewiesen.

Zuletzt kann also gesagt werden, dass die Signalinterpolation die Anforderung erfüllt und teilweise auch übertroffen hat, die Leistung des Spracherkennungssystems nicht herabzusetzen. Ihre volle Wirkung wird sie wahrscheinlich aber erst entfalten können, sobald zu interpolierende Kanäle automatisch erkannt werden können.

# A. Anhang

Die folgende Tabelle T.1 zeigt WER aus den Berechnungen mit dem BioKIT-Erkenner unter Verwendung der PCA in der Vorverarbeitung:

| Interpolation/Speaker/Session | Naiv(4)/551/171 | Naiv(8)/551/171 | Gewichtet(4)/551/171 | Gewichtet(8)/551/171 | Naiv(4)/601/063 | Naiv(8)/601/063 | Gewichtet(4)/601/063 | Gewichtet(8)/601/063 |
|---|---|---|---|---|---|---|---|---|
| Interpolierter Kanal | Referenz: 69,70 | Referenz: 69,70 | Referenz: 69,70 | Referenz: 69,70 | Referenz: 64,65 | Referenz: 64,65 | Referenz: 64,65 | Referenz: 64,65 |
| 1 | 74,75 | 76,77 | 69,70 | 69,70 | 64,65 | 64,65 | 65,66 | 68,69 |
| 2 | 73,74 | 74,75 | 73,74 | 73,74 | 67,68 | 67,68 | 63,64 | 65,66 |
| 3 | 73,74 | 74,75 | 69,70 | 69,70 | 64,65 | 64,65 | 63,64 | 64,65 |
| 4 | 73,74 | 73,74 | 75,76 | 76,77 | 68,69 | 68,69 | 64,65 | 64,65 |
| 5 | 73,74 | 72,73 | 67,68 | 66,67 | 59,60 | 63,64 | 63,64 | 65,66 |
| 6 | 75,76 | 73,74 | 74,75 | 67,68 | 68,69 | 68,69 | 63,64 | 64,65 |
| 7 | 73,74 | 74,75 | 68,69 | 68,69 | 62,63 | 64,65 | 63,64 | 63,64 |
| 8 | 73,74 | 74,75 | 66,67 | 66,67 | 66,67 | 66,67 | 64,65 | 65,66 |
| 9 | 75,76 | 74,75 | 75,76 | 75,76 | 63,64 | 65,66 | 65,66 | 65,66 |
| 10 | 74,75 | 74,75 | 74,75 | 70,71 | 67,68 | 66,67 | 68,69 | 62,63 |
| 11 | 78,79 | 71,72 | 68,69 | 68,69 | 69,70 | 70,71 | 69,70 | 69,70 |
| 12 | 75,76 | 76,77 | 71,72 | 71,72 | 68,69 | 68,69 | 63,64 | 63,64 |
| 13 | 74,75 | 70,71 | 71,72 | 71,72 | 76,77 | 71,72 | 72,73 | 66,67 |
| 14 | 74,75 | 69,70 | 72,73 | 71,72 | 77,78 | 78,79 | 68,69 | 73,74 |
| 15 | 78,79 | 76,77 | 74,75 | 73,74 | 75,76 | 71,72 | 74,75 | 68,69 |
| 16 | 72,73 | 69,70 | 64,65 | 63,64 | 71,72 | 72,73 | 70,71 | 70,71 |
| 17 | 77,78 | 67,68 | 64,65 | 63,64 | 68,69 | 68,69 | 69,70 | 69,70 |
| 18 | 73,74 | 70,71 | 68,69 | 68,69 | 68,69 | 68,69 | 74,75 | 69,70 |
| 19 | 66,67 | 69,70 | 74,75 | 71,72 | 66,67 | 66,67 | 66,67 | 68,69 |
| 20 | 64,65 | 69,70 | 72,73 | 72,73 | 77,78 | 74,75 | 76,77 | 65,66 |
| 21 | 67,68 | 73,74 | 75,76 | 74,75 | 72,73 | 73,74 | 68,69 | 69,70 |
| 22 | 71,72 | 76,77 | 70,71 | 71,72 | 75,76 | 76,77 | 67,68 | 72,73 |
| 23 | 76,77 | 79,80 | 77,78 | 77,78 | 69,70 | 71,72 | 73,74 | 68,69 |
| 24 | 69,70 | 75,76 | 74,75 | 72,73 | 73,74 | 71,72 | 72,73 | 74,75 |
| 25 | 78,79 | 74,75 | 59,60 | 61,62 | 66,67 | 66,67 | 63,64 | 65,66 |
| 26 | 76,77 | 72,73 | 69,70 | 69,70 | 63,64 | 63,64 | 65,66 | 65,66 |
| 27 | 78,79 | 71,72 | 65,66 | 65,66 | 66,67 | 67,68 | 64,65 | 68,69 |
| 28 | 75,76 | 77,78 | 72,73 | 72,73 | 67,68 | 67,68 | 65,66 | 65,66 |
| 29 | 79,80 | 73,74 | 74,75 | 74,75 | 69,70 | 70,71 | 66,67 | 66,67 |
| 30 | 71,72 | 65,66 | 72,73 | 69,70 | 77,78 | 77,78 | 75,76 | 66,67 |
| 31 | 78,79 | 83,84 | 72,73 | 72,73 | 73,74 | 70,71 | 72,73 | 72,73 |
| 32 | 82,83 | 82,83 | 71,72 | 72,73 | 75,76 | 75,76 | 72,73 | 70,71 |
| 33 | 76,77 | 80,81 | 64,65 | 69,70 | 66,67 | 66,67 | 69,70 | 65,66 |
| 34 | 74,75 | 78,79 | 65,66 | 65,66 | 66,67 | 66,67 | 65,66 | 65,66 |
| 35 | 74,75 | 75,76 | 69,70 | 69,70 | 69,70 | 65,66 | 64,65 | 65,66 |

| Naiv(4)/601/064 | Naiv(8)/601/064 | Gewichtet(4)/601/064 | Gewichtet(8)/601/064 | Naiv(4)/703/001 | Naiv(8)/703/001 | Gewichtet(4)/703/001 | Gewichtet(8)/703/001 | Naiv(4)/704/002 | Naiv(8)/704/002 | Gewichtet(4)/704/002 | Gewichtet(8)/704/002 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Referenz: 58,59 | Referenz: 58,59 | Referenz: 58,59 | Referenz: 58,59 | Referenz: 84,85 | Referenz: 84,85 | Referenz: 84,85 | Referenz: 84,85 | Referenz: 77,78 | Referenz: 77,78 | Referenz: 77,78 | Referenz: 77,78 |
| 59,60 | 66,67 | 69,70 | 64,65 | 88,89 | 86,87 | 86,87 | 88,89 | 81,82 | 82,83 | 80,81 | 78,79 |
| 70,71 | 70,71 | 66,67 | 59,60 | 86,87 | 86,87 | 85,86 | 84,85 | 85,86 | 86,87 | 77,78 | 79,80 |
| 64,65 | 64,65 | 59,60 | 60,61 | 84,85 | 84,85 | 85,86 | 87,88 | 79,80 | 80,81 | 80,81 | 80,81 |
| 57,58 | 61,62 | 58,59 | 54,55 | 82,83 | 82,83 | 84,85 | 84,85 | 81,82 | 81,82 | 80,81 | 80,81 |
| 54,55 | 55,56 | 63,64 | 60,61 | 85,86 | 83,84 | 86,87 | 85,86 | 78,79 | 78,79 | 80,81 | 80,81 |
| 61,62 | 62,63 | 63,64 | 60,61 | 86,87 | 85,86 | 88,89 | 84,85 | 77,78 | 78,79 | 78,79 | 77,78 |
| 56,57 | 56,57 | 64,65 | 60,61 | 84,85 | 82,83 | 81,82 | 84,85 | 81,82 | 80,81 | 77,78 | 77,78 |
| 66,67 | 66,67 | 70,71 | 65,66 | 85,86 | 85,86 | 86,87 | 84,85 | 77,78 | 77,78 | 77,78 | 78,79 |
| 71,72 | 69,70 | 69,70 | 66,67 | 81,82 | 81,82 | 82,83 | 81,82 | 78,79 | 78,79 | 77,78 | 77,78 |
| 79,80 | 80,81 | 62,63 | 59,60 | 84,85 | 87,88 | 82,83 | 80,81 | 76,77 | 76,77 | 78,79 | 79,80 |
| 76,77 | 77,78 | 67,68 | 67,68 | 84,85 | 86,87 | 87,88 | 83,84 | 88,89 | 88,89 | 78,79 | 87,88 |
| 68,69 | 68,69 | 68,69 | 67,68 | 82,83 | 84,85 | 89,90 | 81,82 | 81,82 | 80,81 | 73,74 | 78,79 |
| 69,70 | 69,70 | 60,61 | 60,61 | 83,84 | 84,85 | 83,84 | 86,87 | 81,82 | 81,82 | 81,82 | 82,83 |
| 82,83 | 82,83 | 70,71 | 70,71 | 84,85 | 85,86 | 84,85 | 83,84 | 77,78 | 78,79 | 77,78 | 78,79 |
| 77,78 | 77,78 | 63,64 | 63,64 | 83,84 | 83,84 | 82,83 | 86,87 | 82,83 | 81,82 | 81,82 | 79,80 |
| 76,77 | 76,77 | 77,78 | 72,73 | 88,89 | 88,89 | 84,85 | 87,88 | 75,76 | 72,73 | 81,82 | 73,74 |
| 79,80 | 79,80 | 78,79 | 74,75 | 83,84 | 84,85 | 84,85 | 80,81 | 80,81 | 82,83 | 77,78 | 77,78 |
| 76,77 | 73,74 | 81,82 | 74,75 | 82,83 | 83,84 | 83,84 | 82,83 | 82,83 | 82,83 | 77,78 | 79,80 |
| 77,78 | 76,77 | 65,66 | 67,68 | 83,84 | 82,83 | 82,83 | 82,83 | 79,80 | 79,80 | 75,76 | 76,77 |
| 69,70 | 68,69 | 66,67 | 66,67 | 84,85 | 84,85 | 84,85 | 84,85 | 80,81 | 80,81 | 76,77 | 77,78 |
| 69,70 | 69,70 | 67,68 | 64,65 | 84,85 | 86,87 | 83,84 | 85,86 | 81,82 | 81,82 | 81,82 | 81,82 |
| 78,79 | 79,80 | 64,65 | 63,64 | 85,86 | 84,85 | 86,87 | 84,85 | 77,78 | 77,78 | 76,77 | 77,78 |
| 71,72 | 71,72 | 71,72 | 72,73 | 82,83 | 83,84 | 82,83 | 83,84 | 77,78 | 77,78 | 77,78 | 77,78 |
| 64,65 | 64,65 | 73,74 | 73,74 | 87,88 | 85,86 | 84,85 | 84,85 | 76,77 | 76,77 | 79,80 | 77,78 |
| 81,82 | 79,80 | 72,73 | 73,74 | 84,85 | 85,86 | 84,85 | 84,85 | 79,80 | 72,73 | 72,73 | 72,73 |
| 77,78 | 77,78 | 65,66 | 65,66 | 84,85 | 83,84 | 84,85 | 83,84 | 78,79 | 78,79 | 72,73 | 72,73 |
| 73,74 | 70,71 | 72,73 | 72,73 | 83,84 | 83,84 | 88,89 | 83,84 | 78,79 | 78,79 | 79,80 | 76,77 |
| 72,73 | 72,73 | 56,57 | 55,56 | 83,84 | 83,84 | 88,89 | 83,84 | 81,82 | 80,81 | 76,77 | 76,77 |
| 67,68 | 75,76 | 72,73 | 67,68 | 83,84 | 84,85 | 84,85 | 88,89 | 81,82 | 85,86 | 79,80 | 73,74 |
| 77,78 | 70,71 | 67,68 | 70,71 | 84,85 | 83,84 | 81,82 | 82,83 | 85,86 | 85,86 | 79,80 | 79,80 |
| 71,72 | 70,71 | 66,67 | 74,75 | 84,85 | 84,85 | 83,84 | 84,85 | 73,74 | 73,74 | 77,78 | 70,71 |
| 76,77 | 76,77 | 72,73 | 66,67 | 84,85 | 84,85 | 84,85 | 84,85 | 74,75 | 73,74 | 73,74 | 74,75 |
| 75,76 | 75,76 | 69,70 | 72,73 | 84,85 | 83,84 | 83,84 | 84,85 | 75,76 | 75,76 | 78,79 | 78,79 |
| 59,60 | 58,59 | 67,68 | 67,68 | 83,84 | 83,84 | 84,85 | 84,85 | 74,75 | 79,80 | 79,80 | 74,75 |
| 73,74 | 72,73 | 71,72 | 71,72 | 83,84 | 85,86 | 84,85 | 86,87 | 75,76 | 79,80 | 79,80 | 77,78 |

| Naiv(4)/705/002 | Naiv(8)/705/002 | Gewichtet(4)/705/002 | Gewichtet(8)/705/002 | Naiv(4)/706/002 | Naiv(8)/706/002 | Gewichtet(4)/706/002 | Gewichtet(8)/706/002 |
|---|---|---|---|---|---|---|---|
| Referenz: 71,72 | Referenz: 71,72 | Referenz: 71,72 | Referenz: 71,72 | Referenz: 82,83 | Referenz: 82,83 | Referenz: 82,83 | Referenz: 82,83 |
| 69,70 | 69,70 | 69,70 | 69,70 | 84,85 | 84,85 | 82,83 | 84,85 |
| 69,70 | 69,70 | 69,70 | 69,70 | 84,85 | 82,83 | 82,83 | 80,81 |
| 69,70 | 69,70 | 69,70 | 68,69 | 82,83 | 82,83 | 81,82 | 80,81 |
| 74,75 | 74,75 | 72,73 | 71,72 | 80,81 | 80,81 | 82,83 | 82,83 |
| 75,76 | 72,73 | 72,73 | 72,73 | 82,83 | 81,82 | 82,83 | 82,83 |
| 77,78 | 78,79 | 72,73 | 72,73 | 80,81 | 81,82 | 81,82 | 81,82 |
| 75,76 | 74,75 | 74,75 | 75,76 | 82,83 | 82,83 | 82,83 | 82,83 |
| 77,78 | 76,77 | 72,73 | 73,74 | 83,84 | 84,85 | 85,86 | 84,85 |
| 72,73 | 72,73 | 71,72 | 70,71 | 82,83 | 84,85 | 79,80 | 83,84 |
| 73,74 | 72,73 | 72,73 | 74,75 | 79,80 | 84,85 | 79,80 | 79,80 |
| 77,78 | 77,78 | 77,78 | 79,80 | 77,78 | 81,82 | 83,84 | 83,84 |
| 78,79 | 78,79 | 77,78 | 79,80 | 84,85 | 78,79 | 84,85 | 84,85 |
| 77,78 | 77,78 | 76,77 | 77,78 | 87,88 | 83,84 | 87,88 | 83,84 |
| 78,79 | 78,79 | 78,79 | 78,79 | 81,82 | 82,83 | 80,81 | 80,81 |
| 77,78 | 76,77 | 76,77 | 77,78 | 81,82 | 83,84 | 83,84 | 83,84 |
| 76,77 | 76,77 | 76,77 | 76,77 | 83,84 | 81,82 | 85,86 | 85,86 |
| 78,79 | 79,80 | 79,80 | 78,79 | 84,85 | 84,85 | 84,85 | 82,83 |
| 81,82 | 81,82 | 81,82 | 81,82 | 86,87 | 85,86 | 84,85 | 84,85 |
| 82,83 | 80,81 | 80,81 | 82,83 | 87,88 | 87,88 | 87,88 | 86,87 |
| 80,81 | 77,78 | 78,79 | 79,80 | 90,91 | 90,91 | 88,89 | 89,90 |
| 83,84 | 81,82 | 79,80 | 82,83 | 84,85 | 90,91 | 83,84 | 83,84 |
| 78,79 | 77,78 | 77,78 | 76,77 | 85,86 | 86,87 | 85,86 | 83,84 |
| 77,78 | 76,77 | 76,77 | 78,79 | 86,87 | 86,87 | 86,87 | 85,86 |
| 77,78 | 76,77 | 77,78 | 78,79 | 84,85 | 85,86 | 84,85 | 84,85 |
| 76,77 | 78,79 | 79,80 | 78,79 | 85,86 | 86,87 | 83,84 | 84,85 |
| 81,82 | 82,83 | 82,83 | 82,83 | 83,84 | 83,84 | 86,87 | 86,87 |
| 79,80 | 80,81 | 82,83 | 80,81 | 81,82 | 81,82 | 82,83 | 85,86 |
| 78,79 | 78,79 | 79,80 | 78,79 | 82,83 | 86,87 | 88,89 | 86,87 |
| 72,73 | 71,72 | 73,74 | 73,74 | 85,86 | 88,89 | 89,90 | 89,90 |
| 78,79 | 78,79 | 78,79 | 77,78 | 84,85 | 82,83 | 84,85 | 84,85 |
| 75,76 | 75,76 | 76,77 | 76,77 | 81,82 | 81,82 | 81,82 | 84,85 |
| 79,80 | 80,81 | 77,78 | 76,77 | 84,85 | 86,87 | 84,85 | 88,89 |
| 78,79 | 79,80 | 79,80 | 77,78 | 83,84 | 84,85 | 84,85 | 84,85 |
| 79,80 | 79,80 | 78,79 | 78,79 | 84,85 | 84,85 | 83,84 | 83,84 |
| 78,79 | 75,76 | 74,75 | 75,76 | 84,85 | 84,85 | 84,85 | 84,85 |

Die folgende Tabelle T.2 zeigt WER aus den Berechnungen mit dem Janus-Erkenner ohne PCA in der Vorverarbeitung:

| Interpolation/Speaker/Session | Naiv(4)/551/171 | Naiv(8)/551/171 | Gewichtet(4)/551/171 | Gewichtet(8)/551/171 | Naiv(4)/601/063 | Naiv(8)/601/063 | Gewichtet(4)/601/063 | Gewichtet(8)/601/063 |
|---|---|---|---|---|---|---|---|---|
| Interpolierter Kanal | Referenz:41,4 | Referenz:41,4 | Referenz:41,4 | Referenz:41,4 | Referenz:47,5 | Referenz:47,5 | Referenz:47,5 | Referenz:47,5 |
| 1 | 41,4 | 41,4 | 57,6 | 56,6 | 47,5 | 47,5 | 97 | 97 |
| 2 | 34,3 | 34,3 | 63,6 | 69,7 | 94,9 | 94,9 | 92,9 | 91,9 |
| 3 | 40,4 | 40,4 | 40,4 | 43,4 | 76,8 | 76,8 | 64,6 | 64,6 |
| 4 | 47,5 | 47,5 | 43,4 | 43,4 | 94,9 | 94,9 | 56,6 | 58,6 |
| 5 | 64,6 | 64,6 | 44,4 | 41,4 | 93,9 | 93,9 | 65,7 | 54,5 |
| 6 | 83,8 | 83,8 | 85,9 | 92,9 | 87,9 | 87,9 | 88,9 | 87,9 |
| 7 | 86,9 | 86,9 | 75,8 | 74,7 | 89,9 | 89,9 | 89,9 | 89,9 |
| 8 | 62,6 | 62,6 | 36,4 | 36,4 | 86,9 | 86,9 | 65,7 | 58,6 |
| 9 | 45,5 | 45,5 | 41,4 | 41,4 | 50,5 | 43,4 | 47,5 | 46,5 |
| 10 | 98 | 96 | 42,4 | 42,4 | 49,5 | 52,5 | 54,5 | 45,5 |
| 11 | 98 | 96 | 38,4 | 38,4 | 49,5 | 49,5 | 45,5 | 44,4 |
| 12 | 44,4 | 41,4 | 41,4 | 40,4 | 48,5 | 47,5 | 49,5 | 48,5 |
| 13 | 36,4 | 37,4 | 35,4 | 41,4 | 51,5 | 47,5 | 53,5 | 47,5 |
| 14 | 36,4 | 36,4 | 40,4 | 38,4 | 50,5 | 44,4 | 53,5 | 54,5 |
| 15 | 36,4 | 38,4 | 36,4 | 36,4 | 52,5 | 54,5 | 60,6 | 51,5 |
| 16 | 32,3 | 34,3 | 37,4 | 38,4 | 54,5 | 47,5 | 58,6 | 51,5 |
| 17 | 37,4 | 35,4 | 38,4 | 35,4 | 50,5 | 56,6 | 50,5 | 57,6 |
| 18 | 53,5 | 50,5 | 49,5 | 48,5 | 40,4 | 45,5 | 45,5 | 39,4 |
| 19 | 63,6 | 73,7 | 66,7 | 57,6 | 52,5 | 62,6 | 60,6 | 60,6 |
| 20 | 49,5 | 40,4 | 32,3 | **34,3** | 49,5 | 44,4 | 43,4 | 48,5 |
| 21 | **31,3** | **32,3** | 39,4 | **34,3** | **38,4** | 41,4 | 48,5 | 39,4 |
| 22 | 37,4 | 40,4 | 41,4 | 39,4 | 42,4 | 42,4 | 44,4 | 45,5 |
| 23 | 40,4 | 43,4 | 45,5 | 43,4 | 45,5 | 42,4 | 51,5 | 44,4 |
| 24 | 36,4 | 98 | 33,3 | 36,4 | 52,5 | 51,5 | 49,5 | 53,5 |
| 25 | 100 | 99 | 39,4 | 42,4 | 46,5 | 50,5 | 45,5 | 53,5 |
| 26 | 100 | 100 | 38,4 | 36,4 | 41,4 | 52,5 | 47,5 | 49,5 |
| 27 | 99 | 93,9 | **30,3** | 34,3 | 41,4 | **40,4** | **38,4** | 41,4 |
| 28 | 40,4 | 43,4 | 43,4 | 42,4 | 40,4 | 42,4 | 43,4 | 41,4 |
| 29 | 37,4 | 36,4 | 42,4 | 42,4 | 44,4 | 44,4 | 46,5 | 43,4 |
| 30 | 48,5 | 48,5 | 46,5 | 39,4 | 40,4 | 46,5 | 42,4 | **38,4** |
| 31 | 37,4 | 42,4 | 38,4 | 42,4 | 42,4 | 42,4 | 44,4 | 44,4 |
| 32 | 37,4 | 34,3 | 32,3 | 35,4 | 43,4 | 46,5 | 48,5 | 46,5 |
| 33 | 100 | 99 | 36,4 | 41,4 | 43,4 | 48,5 | 39,4 | 48,5 |
| 34 | 98 | 69,7 | 33,3 | 34,3 | 43,4 | 49,5 | 45,5 | 48,5 |
| 35 | 40,4 | 40,4 | 39,4 | 42,4 | 45,5 | 47,5 | 44,4 | 46,5 |

| Naiv(4)/601/064 | Naiv(8)/601/064 | Gewichtet(4)/601/064 | Gewichtet(8)/601/064 | Naiv(4)/703/001 | Naiv(8)/703/001 | Gewichtet(4)/703/001 | Gewichtet(8)/703/001 | Naiv(4)/704/002 | Naiv(8)/704/002 | Gewichtet(4)/704/002 | Gewichtet(8)/704/002 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Referenz:42,4 | Referenz:42,4 | Referenz:42,4 | Referenz:42,4 | Referenz:92,9 | Referenz:92,9 | Referenz:92,9 | Referenz:92,9 | Referenz:62,6 | Referenz:62,6 | Referenz:62,6 | Referenz:62,6 |
| 42,4 | 42,4 | 100 | 100 | 92,9 | 91,9 | 90,9 | 90,9 | 62,6 | 62,6 | 70,7 | 71,7 |
| 100 | 100 | 100 | 100 | 88,9 | 92,9 | 91,9 | 89,9 | 61,6 | 61,6 | 71,7 | 73,7 |
| 97 | 97 | 70,7 | 70,7 | 87,9 | 90,9 | 91,9 | 97 | 65,7 | 65,7 | 76,8 | 67,7 |
| 93,9 | 93,9 | 53,5 | 53,5 | 89,9 | 86,9 | 84,8 | 83,8 | 71,7 | 71,7 | 86,9 | 84,8 |
| 96 | 96 | 47,5 | 59,6 | 83,8 | 90,9 | 85,9 | 93,9 | 101 | 101 | 83,8 | 81,8 |
| 72,7 | 72,7 | 70,7 | 74,7 | 91,9 | 91,9 | 90,9 | 82,8 | 107,1 | 107,1 | 104 | 103 |
| 63,6 | 63,6 | 67,7 | 67,7 | 92,9 | 91,9 | 86,9 | 92,9 | 102 | 102 | 102 | 104 |
| 77,8 | 77,8 | 52,5 | 55,6 | 91,9 | 92,9 | 89,9 | 90,9 | 104 | 104 | 65,7 | 71,7 |
| 42,4 | 42,4 | 45,5 | 45,5 | 91,9 | 96 | 93,9 | 93,9 | 99 | 100 | 67,7 | 69,7 |
| 51,5 | 42,4 | 50,5 | 44,4 | 86,9 | 88,9 | 89,9 | 87,9 | 99 | 100 | **58,6** | 58,6 |
| 80,8 | 57,6 | 45,5 | 43,4 | 88,9 | 90,9 | 91,9 | 84,8 | 97 | 98 | 62,6 | 62,6 |
| 48,5 | 45,5 | 47,5 | 48,5 | 85,9 | 93,9 | 88,9 | 88,9 | 103 | 108,1 | 59,6 | 62,6 |
| 46,5 | 48,5 | 46,5 | 48,5 | 88,9 | 90,9 | 86,9 | 90,9 | 67,7 | 64,6 | 65,7 | 71,7 |
| 43,4 | 42,4 | 43,4 | 40,4 | 91,9 | 88,9 | 88,9 | 89,9 | 63,6 | 62,6 | 61,6 | 65,7 |
| 41,4 | 49,5 | 48,5 | 45,5 | 88,9 | 96 | 97 | 90,9 | 57,6 | **58,6** | 64,6 | 60,6 |
| 38,4 | 39,4 | 39,4 | 40,4 | 88,9 | 90,9 | 88,9 | 86,9 | 67,7 | 63,6 | 60,6 | **57,6** |
| 42,4 | 44,4 | 44,4 | 50,5 | 89,9 | 91,9 | 82,8 | 90,9 | 63,6 | 66,7 | 61,6 | 63,6 |
| 66,7 | 44,4 | 49,5 | 45,5 | 90,9 | 93,9 | 91,9 | 96 | 71,7 | 74,7 | 68,7 | 69,7 |
| 43,4 | 50,5 | 44,4 | 44,4 | 88,9 | 90,9 | 89,9 | 88,9 | 68,7 | 67,7 | 70,7 | 66,7 |
| 42,4 | 58,6 | 42,4 | 45,5 | 87,9 | 86,9 | 85,9 | 89,9 | 67,7 | 64,6 | 70,7 | 67,7 |
| 45,5 | 50,5 | 49,5 | 51,5 | 87,9 | 88,9 | 88,9 | 87,9 | 58,6 | 61,6 | 63,6 | 59,6 |
| 51,5 | 51,5 | 42,4 | 50,5 | **82,8** | 91,9 | 85,9 | 89,9 | 67,7 | 62,6 | 64,6 | 64,6 |
| 41,4 | 37,4 | 41,4 | 40,4 | 87,9 | 89,9 | 90,9 | 89,9 | 67,7 | 68,7 | 70,7 | 68,7 |
| 43,4 | 43,4 | 43,4 | 42,4 | 87,9 | 87,9 | 92,9 | 89,9 | 70,7 | 62,6 | 67,7 | 64,6 |
| 45,5 | 44,4 | 48,5 | 45,5 | 90,9 | 92,9 | 93,9 | 93,9 | 64,6 | 74,7 | 75,8 | 69,7 |
| 43,4 | 46,5 | 44,4 | 40,4 | 92,9 | 90,9 | 89,9 | 90,9 | **55,6** | 60,6 | 59,6 | 59,6 |
| 41,4 | 42,4 | 48,5 | 41,4 | 92,9 | 88,9 | 87,9 | 93,9 | 65,7 | 66,7 | 62,6 | 70,7 |
| 43,4 | 45,5 | 46,5 | 47,5 | 101 | 94,9 | 96 | 94,9 | 63,6 | 67,7 | 67,7 | 69,7 |
| 45,5 | 44,4 | 46,5 | 45,5 | 92,9 | 89,9 | 88,9 | 91,9 | 68,7 | 66,7 | 65,7 | 66,7 |
| 40,4 | 43,4 | 41,4 | 40,4 | 96 | 85,9 | 86,9 | 87,9 | 60,6 | 64,6 | 64,6 | 61,6 |
| 43,4 | 43,4 | 43,4 | 43,4 | 97 | 87,9 | 93,9 | 89,9 | 60,6 | 60,6 | 63,6 | 66,7 |
| 42,4 | 41,4 | 43,4 | 42,4 | 99 | 89,9 | 87,9 | 89,9 | 58,6 | 62,6 | 60,6 | 63,6 |
| 43,4 | 42,4 | **38,4** | 42,4 | 104 | 88,9 | 90,9 | 90,9 | 78,8 | 77,8 | 74,7 | 72,7 |
| 39,4 | 41,4 | 43,4 | **39,4** | 97 | 94,9 | 85,9 | 89,9 | 76,8 | 76,8 | 74,7 | 76,8 |
| **38,4** | **37,4** | 50,5 | 40,4 | 97 | **85,9** | **76,8** | **81,8** | 64,6 | 63,6 | 72,7 | 70,7 |

Die folgende Tabelle T.3 zeigt WER aus den Berechnungen mit dem BioKIT-Erkenner unter Verwendung der PCA in der Vorverarbeitung:

| Interpolation/Speaker/Session | Naiv(4)/551/171 | Naiv(8)/551/171 | Gewichtet(4)/551/171 | Gewichtet(8)/551/171 | Naiv(4)/601/063 | Naiv(8)/601/063 | Gewichtet(4)/601/063 | Gewichtet(8)/601/063 |
|---|---|---|---|---|---|---|---|---|
| Interpolierter Kanal | Referenz:32,3 | Referenz:32,3 | Referenz:32,3 | Referenz:32,3 | Referenz:26,3 | Referenz:26,3 | Referenz:26,3 | Referenz:26,3 |
| 1 | 32,3 | 32,3 | **30,3** | 33,3 | 26,3 | 26,3 | 45,5 | 44,4 |
| 2 | 39,4 | 39,4 | 42,4 | 44,4 | 45,5 | 45,5 | 44,4 | 45,5 |
| 3 | 43,4 | 43,4 | 47,5 | 41,4 | 39,4 | 39,4 | 36,4 | 37,4 |
| 4 | 40,4 | 40,4 | 37,4 | 35,4 | 92,9 | 92,9 | 57,6 | 54,5 |
| 5 | 44,4 | 44,4 | 38,4 | 40,4 | 72,7 | 72,7 | 39,4 | 41,4 |
| 6 | 50,5 | 50,5 | 48,5 | 49,5 | 85,9 | 85,9 | 85,9 | 89,9 |
| 7 | 46,5 | 46,5 | 45,5 | 51,5 | 89,9 | 89,9 | 83,8 | 83,8 |
| 8 | 54,5 | 54,5 | 42,4 | 41,4 | 87,9 | 87,9 | 40,4 | 40,4 |
| 9 | 35,4 | 35,4 | 44,4 | 44,4 | 31,3 | 31,3 | 34,3 | 32,3 |
| 10 | 36,4 | 36,4 | 36,4 | 35,4 | 24,2 | 33,3 | 27,3 | 31,3 |
| 11 | 48,5 | 43,4 | 31,3 | 33,3 | 33,3 | 35,4 | 40,4 | 40,4 |
| 12 | 47,5 | 40,4 | 35,4 | 38,4 | 32,3 | 30,3 | 34,3 | 34,3 |
| 13 | 41,4 | 44,4 | 31,3 | 34,3 | 34,3 | 34,3 | 35,4 | 29,3 |
| 14 | 38,4 | 38,4 | 34,3 | 41,4 | 42,4 | 38,4 | 36,4 | 37,4 |
| 15 | 38,4 | 37,4 | 33,3 | 40,4 | 36,4 | 34,3 | 45,5 | 27,3 |
| 16 | 46,5 | 47,5 | 45,5 | 48,5 | 30,3 | 27,3 | 36,4 | 27,3 |
| 17 | 38,4 | 36,4 | 41,4 | 38,4 | 47,5 | 36,4 | 41,4 | 35,4 |
| 18 | 53,5 | 43,4 | 46,5 | 44,4 | 26,3 | 33,3 | 34,3 | 39,4 |
| 19 | 59,6 | 50,5 | 54,5 | 53,5 | 28,3 | 28,3 | 31,3 | 38,4 |
| 20 | 47,5 | 45,5 | 44,4 | 45,5 | 32,3 | 31,3 | 37,4 | 32,3 |
| 21 | 43,4 | 37,4 | 37,4 | 39,4 | 32,3 | 30,3 | 31,3 | 29,3 |
| 22 | 37,4 | 32,3 | 34,3 | 33,3 | **24,2** | **24,2** | 29,3 | **24,2** |
| 23 | 42,4 | 39,4 | 43,4 | 41,4 | **24,2** | **24,2** | **24,2** | 27,3 |
| 24 | 40,4 | 47,5 | 40,4 | 39,4 | 43,4 | 36,4 | 36,4 | 38,4 |
| 25 | 36,4 | 87,9 | 42,4 | 39,4 | 48,5 | 47,5 | 43,4 | 38,4 |
| 26 | 94,9 | 98 | 40,4 | 38,4 | 48,5 | 47,5 | 48,5 | 44,4 |
| 27 | 98 | 92,9 | 35,4 | 37,4 | 27,3 | 28,3 | 26,3 | 32,3 |
| 28 | 41,4 | 46,5 | 45,5 | 42,4 | 39,4 | 33,3 | 36,4 | 34,3 |
| 29 | 41,4 | 36,4 | 47,5 | 36,4 | 34,3 | 32,3 | 33,3 | 31,3 |
| 30 | 34,3 | 34,3 | 33,3 | 32,3 | 33,3 | 31,3 | 33,3 | 31,3 |
| 31 | 41,4 | 40,4 | 43,4 | 38,4 | 33,3 | 33,3 | 27,3 | 29,3 |
| 32 | 39,4 | 39,4 | 39,4 | 45,5 | 26,3 | 26,3 | 31,3 | 26,3 |
| 33 | 84,8 | 57,6 | 31,3 | 31,3 | 36,4 | 33,3 | 36,4 | 32,3 |
| 34 | 97 | 65,7 | 41,4 | 39,4 | 26,3 | 26,3 | 32,3 | 26,3 |
| 35 | 42,4 | 41,4 | 39,4 | 38,4 | 32,3 | 25,3 | 25,3 | 30,3 |

| Naiv(4)/601/064 | Naiv(8)/601/064 | Gewichtet(4)/601/064 | Gewichtet(8)/601/064 | Naiv(4)/703/001 | Naiv(8)/703/001 | Gewichtet(4)/703/001 | Gewichtet(8)/703/001 | Naiv(4)/704/002 | Naiv(8)/704/002 | Gewichtet(4)/704/002 | Gewichtet(8)/704/002 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Referenz:29,3 | Referenz:29,3 | Referenz:29,3 | Referenz:29,3 | Referenz:74,7 | Referenz:74,7 | Referenz:74,7 | Referenz:74,7 | Referenz:59,6 | Referenz:59,6 | Referenz:59,6 | Referenz:59,6 |
| **29,3** | 29,3 | 98 | 98 | 74,7 | 74,7 | 75,8 | 75,8 | 59,6 | 59,6 | 73,7 | 75,8 |
| 98 | 98 | 98 | 98 | 74,7 | 74,7 | 76,8 | 78,8 | 59,6 | 59,6 | 62,6 | 60,6 |
| 96 | 96 | 63,6 | 64,6 | 74,7 | 74,7 | 76,8 | 78,8 | 59,6 | 59,6 | **54,5** | 66,7 |
| 58,6 | 58,6 | 30,3 | **26,3** | 73,7 | 73,7 | 75,8 | 75,8 | 61,6 | 61,6 | 55,6 | 62,6 |
| 37,4 | 37,4 | 34,3 | 36,4 | 76,8 | 76,8 | 73,7 | 74,7 | 77,8 | 77,8 | 64,6 | 62,6 |
| 40,4 | 40,4 | 49,5 | 48,5 | 79,8 | 79,8 | 77,8 | 79,8 | 98 | 98 | 92,9 | 93,9 |
| 37,4 | 37,4 | 37,4 | 37,4 | 71,7 | 71,7 | 74,7 | **66,7** | 98 | 98 | 98 | 96 |
| 38,4 | 38,4 | 38,4 | 38,4 | 71,7 | 71,7 | 74,7 | **66,7** | 92,9 | 92,9 | 61,6 | 64,6 |
| 34,3 | 34,3 | 35,4 | 35,4 | 80,8 | 80,8 | 81,8 | 81,8 | 59,6 | 59,6 | 62,6 | 70,7 |
| 32,3 | 28,3 | 34,3 | 31,3 | 79,8 | 79,8 | 81,8 | 76,8 | 57,6 | 54,5 | 58,6 | 58,6 |
| 67,7 | 48,5 | 33,3 | 34,3 | 72,7 | 72,7 | 81,8 | 78,8 | 65,7 | 64,6 | 58,6 | 58,6 |
| 31,3 | 34,3 | 32,3 | 31,3 | 92,9 | 82,8 | 84,8 | 82,8 | 68,7 | 63,6 | 65,7 | 66,7 |
| 37,4 | 37,4 | 38,4 | 37,4 | 79,8 | 79,8 | 78,8 | 79,8 | 62,6 | 65,7 | 62,6 | 62,6 |
| 32,3 | 35,4 | 36,4 | 33,3 | 83,8 | 79,8 | 84,8 | 80,8 | 62,6 | 63,6 | 68,7 | 62,6 |
| 40,4 | 43,4 | 45,5 | 42,4 | 82,8 | 77,8 | 83,8 | 80,8 | 63,6 | 60,6 | 66,7 | 67,7 |
| 32,3 | 32,3 | 37,4 | 34,3 | 74,7 | 79,8 | **73,7** | 74,7 | 63,6 | 63,6 | 63,6 | 62,6 |
| 32,3 | 33,3 | 33,3 | 34,3 | **67,7** | 74,7 | 79,8 | 75,8 | 66,7 | 61,6 | 62,6 | 65,7 |
| 59,6 | 45,5 | 33,3 | 32,3 | 80,8 | 81,8 | 75,8 | 81,8 | 64,6 | 63,6 | 64,6 | 62,6 |
| 34,3 | 36,4 | 32,3 | 38,4 | 89,9 | 86,9 | 92,9 | 93,9 | 63,6 | 63,6 | 67,7 | 58,6 |
| 36,4 | 41,4 | 42,4 | 32,3 | 76,8 | 84,8 | 87,9 | 80,8 | **51,5** | **51,5** | **54,5** | 57,6 |
| **29,3** | 36,4 | 41,4 | 35,4 | 80,8 | 76,8 | 82,8 | 80,8 | 63,6 | 61,6 | 61,6 | 62,6 |
| 37,4 | 42,4 | 35,4 | 34,3 | 73,7 | **71,7** | 75,8 | 76,8 | 61,6 | 62,6 | 63,6 | 60,6 |
| 32,3 | **27,3** | 30,3 | 29,3 | 78,8 | 76,8 | 76,8 | 76,8 | 65,7 | 56,6 | 59,6 | **52,5** |
| 29,3 | 33,3 | **28,3** | 30,3 | 76,8 | 78,8 | 77,8 | 79,8 | 68,7 | 64,6 | 60,6 | 62,6 |
| 33,3 | 30,3 | 36,4 | 32,3 | 78,8 | 77,8 | 75,8 | 80,8 | 62,6 | 62,6 | 63,6 | 59,6 |
| 32,3 | 34,3 | 41,4 | 31,3 | 81,8 | 79,8 | 80,8 | 80,8 | 59,6 | 61,6 | 56,6 | 60,6 |
| 32,3 | 36,4 | 36,4 | 33,3 | 79,8 | 74,7 | 77,8 | 78,8 | 67,7 | 62,6 | 62,6 | 62,6 |
| 39,4 | 38,4 | 36,4 | 39,4 | 77,8 | 76,8 | 77,8 | 80,8 | 67,7 | 61,6 | 65,7 | 67,7 |
| 32,3 | 31,3 | 32,3 | 33,3 | 84,8 | 85,9 | 85,9 | 81,8 | 64,6 | 59,6 | 64,6 | 65,7 |
| 31,3 | 27,3 | 33,3 | 31,3 | 82,8 | 81,8 | 86,9 | 83,8 | 66,7 | 69,7 | 64,6 | 63,6 |
| 32,3 | 32,3 | 31,3 | 35,4 | 87,9 | 88,9 | 90,9 | 89,9 | 61,6 | 65,7 | 60,6 | 63,6 |
| 35,4 | 33,3 | 33,3 | 33,3 | 86,9 | 90,9 | 88,9 | 86,9 | 67,7 | 60,6 | 61,6 | 63,6 |
| 33,3 | 34,3 | 33,3 | 33,3 | 81,8 | 81,8 | 81,8 | 79,8 | 59,6 | 60,6 | 64,6 | 59,6 |
| 32,3 | 33,3 | 32,3 | 34,3 | 74,7 | 77,8 | 80,8 | 74,7 | 74,7 | 61,6 | 75,8 | 62,6 |
| 32,3 | 33,3 | 40,4 | 38,4 | 83,8 | 85,9 | 81,8 | 81,8 | 65,7 | 60,6 | 71,7 | 66,7 |

| Interpolations-methode | SpeakerID, SessionID | Mittelwert (mit PCA) | Minimum (mit PCA) |
|---|---|---|---|
| Naiv(4) | 551, 171 | 74,78% | 64,65% |
| Naiv(8) | 551, 171 | 74,37% | 65,66% |
| Gewichtet(4) | 551, 171 | 70,88% | 59,60% |
| Gewichtet(8) | 551, 171 | 70,45% | 61,62% |
| Naiv(4) | 601, 063 | 69,52% | 59,60% |
| Naiv(8) | 601, 063 | 69,38% | 63,64% |
| Gewichtet(4) | 601, 063 | 68,17% | 63,64% |
| Gewichtet(8) | 601, 063 | 67,53% | 62,63% |
| Naiv(4) | 601, 064 | 71,26% | 54,55% |
| Naiv(8) | 601, 064 | 71,31% | 55,56% |
| Gewichtet(4) | 601, 064 | 68,17% | 56,57% |
| Gewichtet(8) | 601, 064 | 66,70% | 54,55% |
| Naiv(4) | 703, 001 | 84,79% | 81,82% |
| Naiv(8) | 703, 001 | 84,91% | 81,82% |
| Gewichtet(4) | 703, 001 | 85,11% | 81,82% |
| Gewichtet(8) | 703, 001 | 84,73% | 81,81% |
| Naiv(4) | 704, 002 | 79,71% | 73,74% |
| Naiv(8) | 704, 002 | 79,86% | 72,73% |
| Gewichtet(4) | 002, 171 | 78,38% | 72,73% |
| Gewichtet(8) | 002, 171 | 78,04% | 70,71% |
| Naiv(4) | 705, 001 | 77,40% | 69,70% |
| Naiv(8) | 705, 001 | 77,03% | 69,70% |
| Gewichtet(4) | 705, 001 | 76,65% | 69,70% |
| Gewichtet(8) | 705, 001 | 76,85% | 68,69% |
| Naiv(4) | 706, 001 | 84,07% | 77,78% |
| Naiv(8) | 706, 001 | 84,56% | 78,79% |
| Gewichtet(4) | 706, 001 | 84,44% | 79,80% |
| Gewichtet(8) | 706, 001 | 84,50% | 79,80% |

Abbildung A.1: Wortfehlerraten und Minima der einzelnen Sessions mit Interpolation für den BioKIT-Erkenner mit PCA.

| Interpolations-methode | SpeakerID, SessionID | Mittelwert (onhe PCA) | Mittelwert (mit PCA) | Minimum (ohne PCA) | Minimum (mit PCA) |
|---|---|---|---|---|---|
| Naiv(4) | 551, 171 | 56,48% | 48,51% | 31,3% | 32,30% |
| Naiv(8) | 551, 171 | 57,37% | 47,07% | 32,3% | 32,30% |
| Gewichtet(4) | 551, 171 | 42,77% | 39,31% | 32,3% | 31,30% |
| Gewichtet(8) | 551, 171 | 43,95% | 39,33% | 34,3% | 32,30% |
| Naiv(4) | 601, 063 | 54,97% | 38,4% | 41,41% | 24,20% |
| Naiv(8) | 601, 063 | 55,93% | 40,4% | 40,41% | 24,20% |
| Gewichtet(4) | 601, 063 | 54,98% | 38,4% | 38,94% | 24,20% |
| Gewichtet(8) | 601, 063 | 53,82% | 38,4% | 37,66% | 24,20% |
| Naiv(4) | 601, 064 | 53,70% | 38,4% | 40,03% | 29,30% |
| Naiv(8) | 601, 064 | 53,27% | 37,4% | 39,71% | 27,30% |
| Gewichtet(4) | 601, 064 | 50,91% | 38,4% | 40,17% | 28,30% |
| Gewichtet(8) | 601, 064 | 50,79% | 39,4% | 39,13% | 26,30% |
| Naiv(4) | 703, 001 | 91,34% | 82,8% | 79,18% | 67,70% |
| Naiv(8) | 703, 001 | 90,85% | 85,9% | 79,01% | 71,70% |
| Gewichtet(4) | 703, 001 | 89,35% | 80,46% | 80,46% | 73,70% |
| Gewichtet(8) | 703, 001 | 90,15% | 79,42% | 79,42% | 66,70% |
| Naiv(4) | 704, 002 | 73,85% | 55,6% | 66,72% | 51,50% |
| Naiv(8) | 704, 002 | 74,37% | 58,6% | 65,01% | 51,50% |
| Gewichtet(4) | 002, 171 | 69,81% | 58,6% | 65,07% | 54,50% |
| Gewichtet(8) | 002, 171 | 69,75% | 57,6% | 64,90% | 52,50% |

Abbildung A.2: Mittelwerte und Minima aus Berechnungen des Janus-Erkenners mit und ohne PCA in der Vorverarbeitung

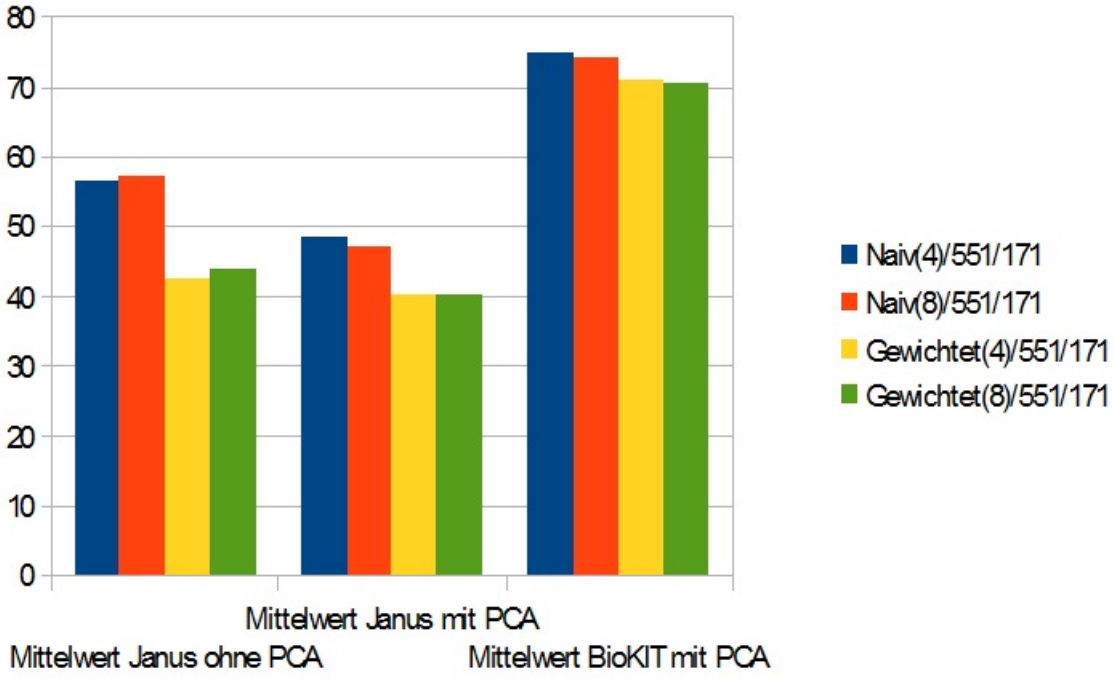


Abbildung A.3: Mittelwerte von Sprecher 551, Aufnahme 171

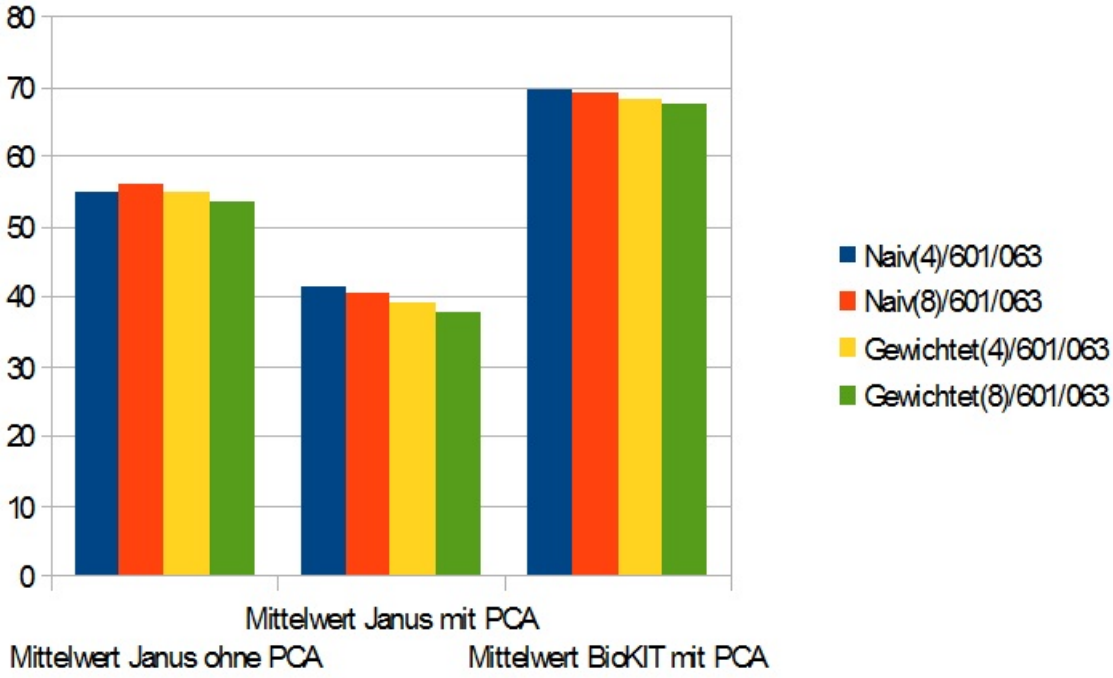


Abbildung A.4: Mittelwerte von Sprecher 601, Aufnahme 063

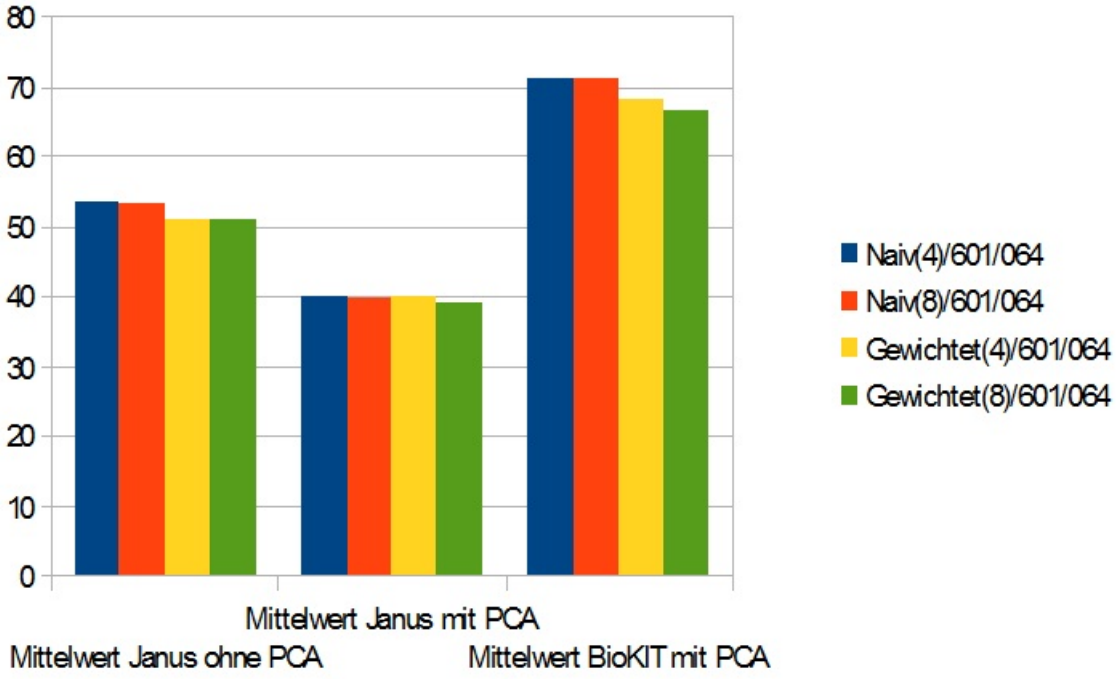


Abbildung A.5: Mittelwerte von Sprecher 601, Aufnahme 064

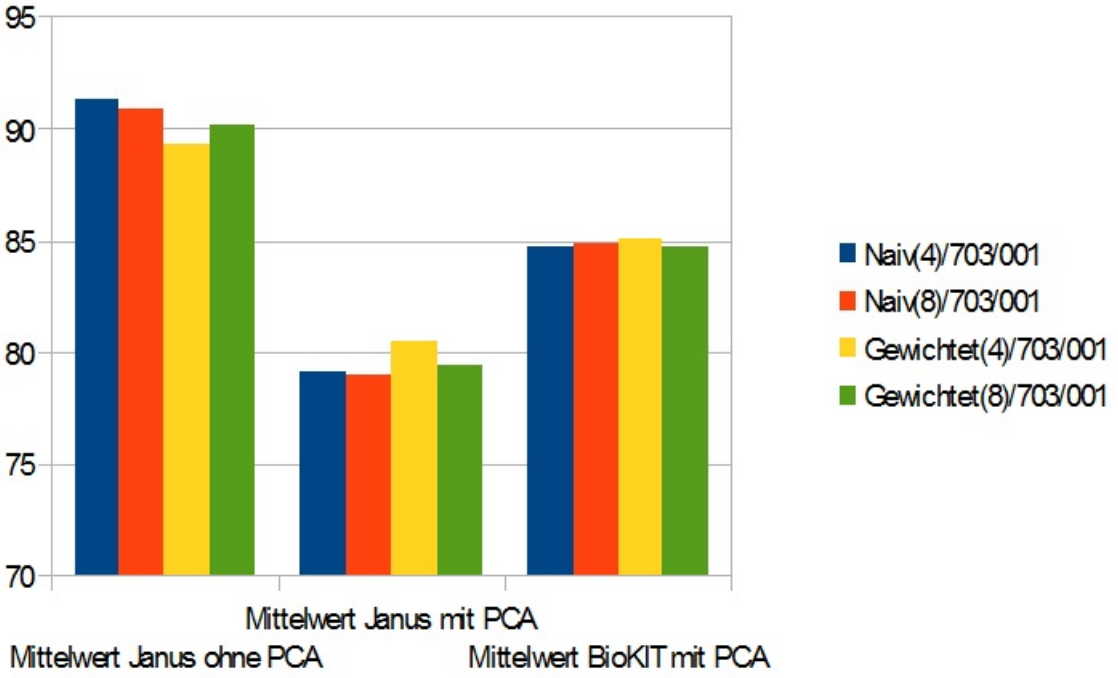


Abbildung A.6: Mittelwerte von Sprecher 703, Aufnahme 001

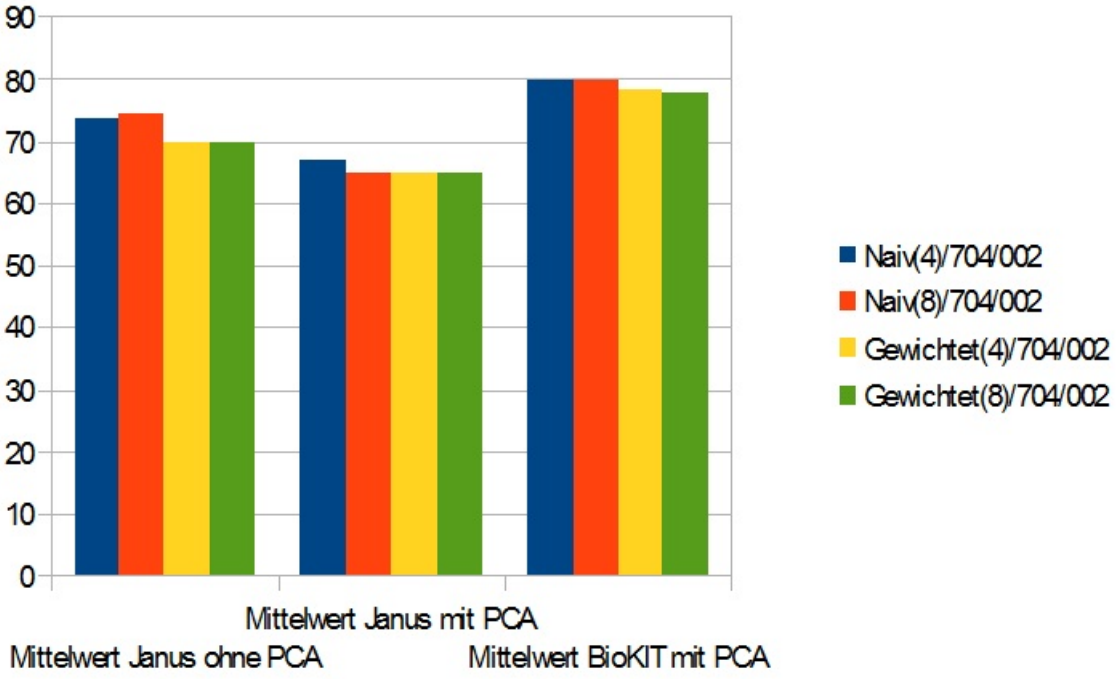


Abbildung A.7: Mittelwerte von Sprecher 704, Aufnahme 002

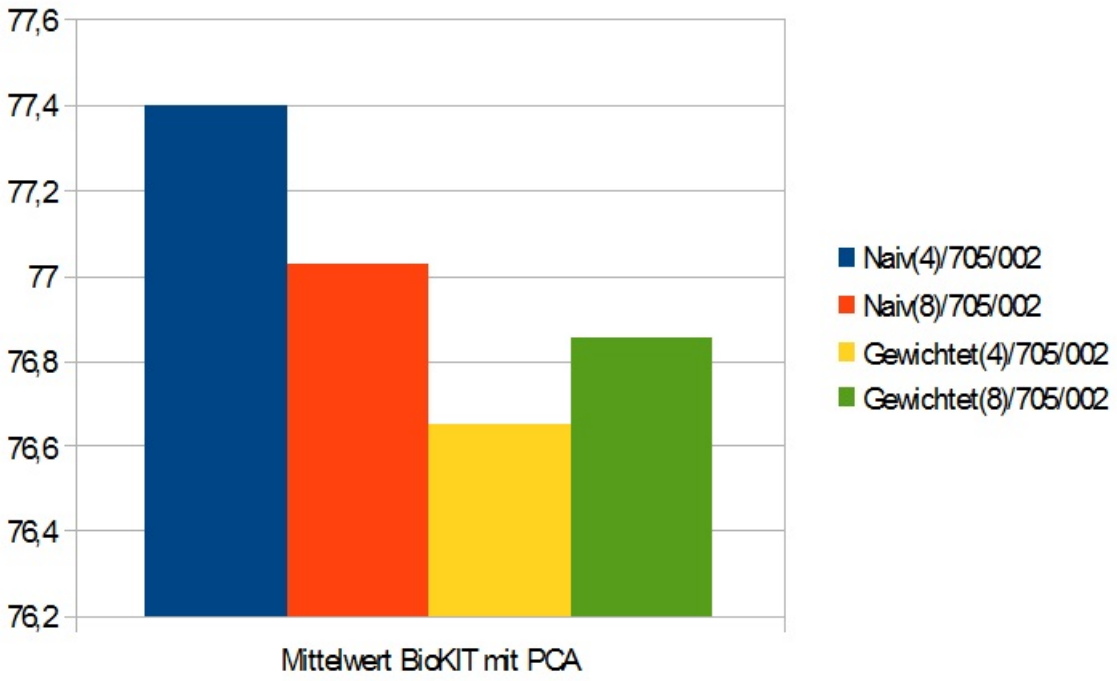


Abbildung A.8: Mittelwerte von Sprecher 705, Aufnahme 001

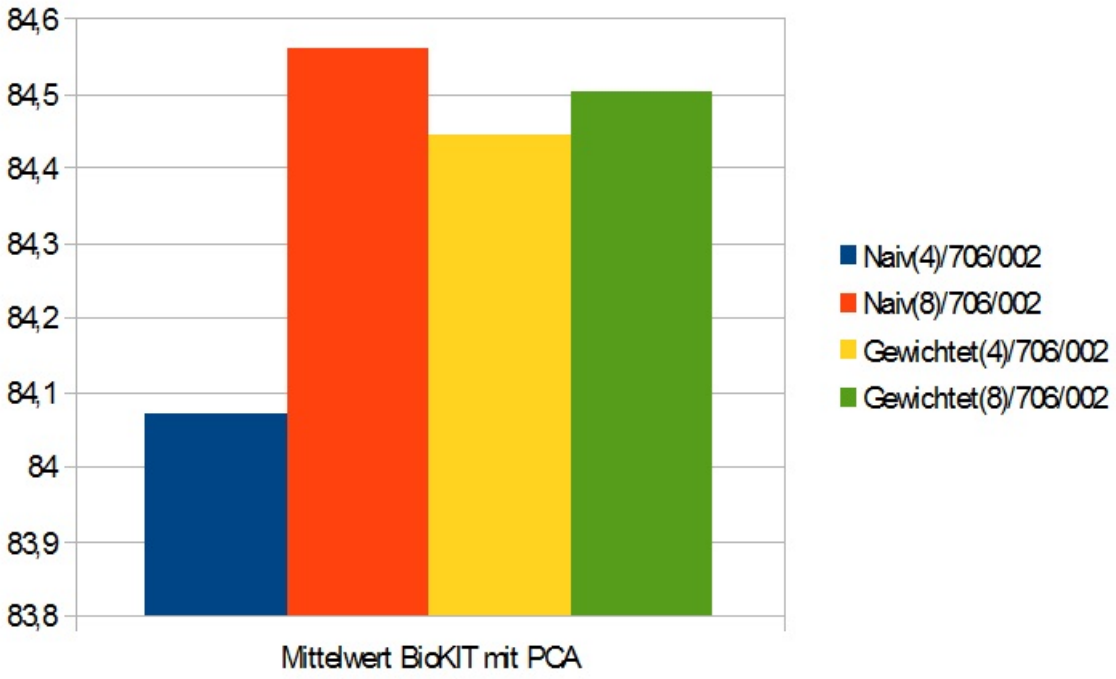


Abbildung A.9: Mittelwerte von Sprecher 706, Aufnahme 001

# Literaturverzeichnis


[1] *Bertelsmann Lexikon.* Bertelsmann Lexikon Verlag, Gütersloh, 2001.

[2] Denby, Bruce, Tanja Schultz, Kiyoshi Honda, Thomas Hueber und James Gilbert: *Silent Speech Interfaces.* Speech Communication, 52(4):270 – 287, 2010.

[3] Heger, Dominic, Felix Putze, Christoph Amma, Thomas Wielatt, Igor Plotkin, Michael Wand und Tanja Schultz: *BiosignalsStudio: A flexible Framework for Biosignal Capturing and Processing.* In: *Proc. KI*, 2010.

[4] Hoole, Philip: *Einsatz der elektromagnetischen Artikulographie bei der Analyse lingualer Sprechbewegungen.* V. Hahn, C. Schneider & H. Hahn (eds.) Schauplatz Mund. Das orofaziale System als sensomotorische Einheit, Seiten 101 – 114, 1999.

[5] Hueber, Thomas, Elie-Laurent Benaroya, Gérard Chollet, Bruce Denby, Gérard Dreyfus und Maureen Stone: *Development of a Silent Speech Interface Driven by Ultrasound and Optical Images of the Tongue and Lips.* Speech Communication, 52:288 – 300, 2010.

[6] Hueber, Thomas, Gérard Chollet, Bruce Denby, Maureen Stone und Leila Zouari: *Ouisper: Corpus Based Synthesis Driven by Articulatory Data.* In: *Proceedings of the 16th International Congress of Phonetic Sciences*, 2007.

[7] Janke, Matthias: *Spektrale Methoden zur EMG-basierten Erkennung lautloser Sprache.* Diplomarbeit, Universität Karlsruhe (TH), Germany, 2010.

[8] Maier-Hein, Lena: *Speech Recognition Using Surface Electromyography.* Diplomarbeit, Universität Karlsruhe (TH), Germany, 2005.

[9] OTBiolettronica: *Arrays & Matrixes pin out. http://www.otbioelettronica.it.*

[10] OTBiolettronica: *User manuel EMG-USB2 Version 2.*

[11] Schulte, Christopher: *Aufbau eines EMG-basierten Spracherkennungssystems unter Verwendung von Elektrodenarrays.* Bachelorarbeit, Karlsruher Institut für Technologie (KIT), Germany, 2011.

[12] Schultz, Tanja und Michael Wand: *Modeling Coarticulation in Large Vocabulary EMG-based Speech Recognition.* Speech Communication, 52(4):341 – 353, 2010.

[13] Wester, Marek: *Unspoken Speech - Speech Recognition Based On Electroencephalography.* Diplomarbeit, Lehrstuhl Prof. Waibel Interactive Systems Laboratories Carnegie Mellon University, Pittsburgh, PA, USA and Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany, 2006.