HSR Hochschule für Technik
Rapperswil

Studienarbeit

Studie zur Erkennung von Phonemen

Autoren:
Claude Baumann
Rolf Schönenberger

Betreuer:
Prof. Dr. Ruedi Stoop

21. Dezember 2011

Abstract

Für die Interaktion zwischen Mensch und Maschine stellt die Erkennung von Phonemen ein fundamentales Problem dar. Trotz grosser Anstrengungen sind auch heute nur verhältnismässig schlechte Verfahren bekannt. Heutige Methoden zur Spracherkennung funktionieren nur, weil der Alltags-Wortschatz stark beschränkt wird und viele Übergänge zwischen Phonemen äusserst unwahrscheinlich sind. Eine statistische Vorgehensweise mittels Hidden Markov Models führt dann zur gewünschten Zuverlässigkeit der Spracherkennung, obwohl die einzelnen Phoneme schlecht erkannt werden [1, 2].
In unserer Studie untersuchen wir, wieso die Erkennung von Phonemen dem Computer so grosse Schwierigkeiten bereitet. Wir untersuchen in geeigneten Räumen der Lautdarstellung, wo sich darin die Träger der Phoneme befinden. Bezüglich der Phoneme setzen wir das Hauptaugenmerk auf die Familie der Frikative, da diese die schlechtesten Ergebnisse bei der Erkennung liefern [3].
Als Darstellungsraum haben wir einführend den Fourierraum mit dem Raum der Wavelets als Basis verglichen. Wir haben festgestellt, dass sich in der Fourierdarstellung durchaus gewisse Muster erkennen lassen, welche für das Phonem charakteristisch sind. Diese reichen allerdings für eine Klassifizierung nicht aus, weshalb die Waveletdarstellungen vorzuziehen sind.
Als Klassifizierungsverfahren haben wir alternativ Clustering und Ansätze einfacher neuronaler Netze verfolgt. Als einfaches Neuron haben wir das sogenannte holographische Neuron verwendet, welches in den Räumen der komplexen Zahlen arbeitet [4]. Obwohl es für seine Effizienz bekannt ist, konnte es auch nach Tagen des Lernens nicht mehr als eine durchschnittliche Erkennungsrate von rund 20 Prozent erreichen.
Während herkömmliche Clusteringverfahren wie k-Means daran scheitern, korrekte Cluster für Phoneme zu finden oder unbekannte Phoneme einzuordnen, erreicht das Hebbian Learning Clustering Verfahren dies mit einer durchschnittlichen Erkennungsrate von mehr als 60 Prozent, ohne dass dabei zu grosse Phonem-Gruppen gebildet werden müssen [5]. Damit wurden vergleichbare Ansätze der Literatur wesentlich übertroffen [6, 7, 8], was ein Ziel dieser Arbeit war. Die Erkennungsrate kann dabei durch eine Verbesserung der Waveletdarstellung optimiert werden.

Erklärung

Hiermit erklären wir, dass wir die vorliegende Arbeit selber und ohne fremde Hilfe durchgeführt haben, ausser derjenigen, welche explizit in der Aufgabenstellung erwähnt ist oder mit dem Betreuer schriftlich vereinbart wurde, sowie sämtliche verwendeten Quellen erwähnt und gemäss gängigen wissenschaftlichen Zitierregeln korrekt angegeben haben.

Rapperswil, den 21. Dezember 2011

Claude Baumann:

Rolf Schönenberger:

Inhaltsverzeichnis

Kapitel 1

Einführung

1.1 Motivation

Die Unterscheidung von stimmlosen Frikativen ist noch heute ein grosses Problem. Die direkte Erkennungsrate von Phonemen liegt unter 50 Prozent, wobei sie noch wesentlich von den Grössen der gewählten Phonemklassen abhängt [6, 7, 8]. In dieser Studie werden verschiedene Verfahren und Darstellungsformen untersucht, um starke Eigenschaften in den jeweiligen Phonemen zu finden. Zur Auswahl stehen der Fourierraum, Wavelets und Local Cosine Packages. Letztere wurden aus Zeitgründen in dieser Studie nicht betrachtet.

1.2 Datenbasis

Die Arbeit zur Erkennung der Phoneme basiert auf dem Datenset des TIMIT Acoustic-Phonetic Continuous Speech Corpus [9]. Verwendet werden in dieser Studie die männlichen Sprecher der Sprachregionen DR1-DR5. Die TIMIT Sprachdaten haben eine Sampling-Rate von 16kHz und sind in zwei Blöcke unterteilt. Zwei Drittel der Daten bilden das Lernset und ein Drittel der Daten das Testset. Die Lern- und Testdaten sind sowohl vom Sprecher als auch von den gesprochenen Sätzen unabhängig.
In unserer Arbeit wird aus Zeitgründen das Hauptaugenmerk nur auf die stimmlosen Frikative s, z, sh, ch, f und dh gelegt. Aus Interesse wurde später noch das Phonem r einbezogen.

Kapitel 2

Darstellungsräume

2.1 Fourierraum

Mit diesem Ansatz versucht man, das Eingangssignal in seine Frequenzspektren zu unterteilen und die jeweiligen Koeffizienten zu vergleichen, um eindeutige Merkmale zu finden. Schneidet man alle Koeffizienten unterhalb eines individuellen Schwellwertes ab, werden Lücken im Spektrum sichtbar.

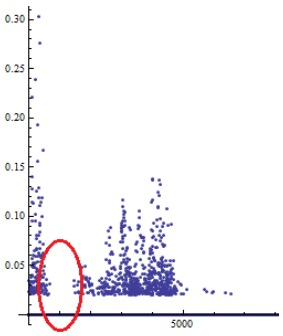

Abbildung 2.1: Überlagerung von Frequenzspektren von 15 *s*-Phonemen (Schwellwert 0.02).

Auf der Abbildung 2.1 ist zu sehen, dass das Phonem *s* im Frequenzbereich 800Hz-1200Hz keine Ausschläge hat. Diese Lücken im Spektrum werden erreicht, nachdem alle Koeffizienten kleiner als 0.02 abgeschnitten werden. Tabelle 2.1 zeigt den Schwellwert und die Frequenzlücken.

Versuche, Testdaten über die erhaltenen Lücken im Spektrum zu klassifizieren schlugen fehl. Es wurden Phoneme an Stellen gefunden, wo andere Phoneme lagen. Beispielsweise wurde das Phonem *s* entdeckt, wo klar das Phonem *r* zu hören

Phonem	Schwellwert	Lücken in Hz
s	0.02	800-1200
t	0.1	1800-2200
f	0.1	2700-2900
a	0.1	keine Lücke

Tabelle 2.1: Phoneme mit Schwellwert und Frequenzlücke.

war. Unsere Erfolgsquote lag bei unter 10 Prozent. Dies wird daran liegen, dass unterschiedliche Phoneme im selben Frequenzbereich Lücken aufweisen. Da die Frequenzen pro Phonem von Sprecher zu Sprecher unterschiedlich sind, werden wohl auch immer andere Lücken gefunden.

2.2 Waveleträume

Die Wavelets sind ein mathematisches Werkzeug zur Darstellung und zur Näherung von Funktionen. Durch fortgesetzte Summation gelangen wir von einer gröberen zu immer feineren Darstellungen [vgl. 10]. Wavelets werden in der Fachliteratur für verschiedenste Anwendungen verwendet, unter anderem auch in der Sprachanalyse. Die meisten Ansätze der Audioanalyse verwenden das Morlet-Wavelet (kontinuierlich), sowie die Wavelets Cohen-Daubechies-Feauveau und Daubechies (diskret). Beide Wavelet Typen, die kontinuierlichen sowie die diskreten Wavelets, lassen beim Betrachten von Phonemen unterscheidbare Muster erkennen.

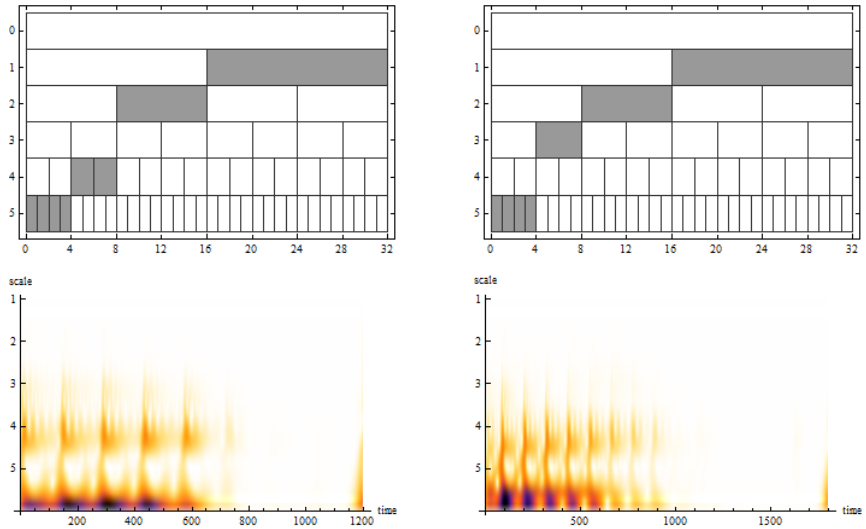

Abbildung 2.2: Vergleich von 2 r-Phonemen. oben: diskretes CDF-Wavelet, unten: kontinuierliches Morlet-Wavelet.

Kapitel 3

Verfahren zur Erkennung

In diesem Kapitel werden zwei Verfahren zur Erkennung von Phonemen untersucht. Zum einen das Clustering in geeigneten Darstellungsräumen und zum anderen die Klassifizierung mittels eines holographischen Neurons.

3.1 Holographisches Neuron

Das holographische Neuron (auch holographische Synapse genannt) ist vor allem auf Grund seiner hohen Verarbeitungsgeschwindigkeit und Effizienz bekannt, dabei besitzt es eine grosse Speicherfähigkeit.
Wie bei klassischen neuronalen Netzen beginnt man mit einem Eingabemuster. Dabei verfolgt man das Ziel, jeden Eingabevektor mit seinem Antwortvektor zu verknüpfen. Den wichtigsten Teil des Holographen stellt eine $m \times n$ Matrix dar. Dabei bezeichnet n die Dimension des Eingabevektors und m die Dimension des Antwortvektors. Anfangs wird die Matrix mit Zufallswerten initialisiert, welche während des Lernprozesses durch die Eingabemuster nach einem Gradientenabstieg verändert werden.
Der Algorithmus verhält sich gemäss $O(nm)$. Der Speicherbedarf ist ebenfalls von der Ordnung mn, wobei dieser nicht durch die Anzahl der assoziierten Muster abhängt [vgl. 4].

Der Algorithmus ist im Skript [4] genau beschrieben.

3.1.1 Lernen der wav-Rohdaten

In einem ersten Schritt wurde versucht, ein holographisches Neuron auf die TIMIT wav-Rohdaten der Phoneme s, sh, z, dh und f zu trainieren. Dabei zeigt sich die Problematik, dass die Phoneme von 400 Datenpunkten (0.025 Sekunden) bis 4000 Datenpunkten (0.25 Sekunden) variieren.

Um dieses Problem zu lösen, werden alternativ folgende Ansätze untersucht:

- Kurze Phoneme erweitern
- Lange Phoneme kürzen

Am einfachsten war es, die kurzen Phoneme mit Füllwerten zu erweitern. Der Vorteil von dieser Methode liegt darin, dass keine Information verloren geht. Dabei stellte sich die Frage, ob mit Nullen oder mit Zufallswerten erweitert werden soll.
Der Holograph war leider in beiden Fällen nicht fähig, mit den erweiterten Daten vernünftig zu arbeiten, was daran liegen kann, dass der Holograph im Raum der komplexen Zahlen operiert und die Informationen in Phasen und nicht in Absolutwerten kodiert werden.
Beim erweitern mit Zufallswerten sollte beachtet werden, dass bei jeder Lernepoche neue Werte gebildet werden, damit nicht Zufallswerte gelernt werden. Die Abbildung 3.1 zeigt den eher schlechten Lernfortschritt des Holographen. Der schlechte Performance Index kann unterschiedliche Gründe haben. Zum einen sind die wav-Rohdaten nicht als Phonemdarstellung geeignet, zum anderen ist der Anteil an Zufallswerten bei kurzen Phonemen viel zu gross.

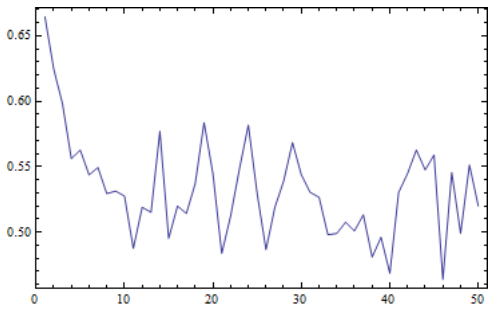

Abbildung 3.1: Performance-Index beim Lernen von Phonemen (Erweitert mit Zufallswerten).

Die Kürzung von langen Phonemen birgt andere Probleme. Hier ist es möglich, dass die Phoneme der TIMIT Datenbank nicht ganz genau getroffen wurden. Diese Tatsache macht es extrem schwierig, zu entscheiden, welcher Teil abgeschnitten werden kann, ohne zu viele Informationen zu verlieren. Vor allem kann nicht entschieden werden, welcher Teil für das Phonem charakteristisch ist und welcher nicht.

3.1.2 Lernen der Diskreten Wavelet Koeffizienten

Die diskreten Wavelets als Darstellungsform der Phoneme haben den Vorteil, immer die gleiche Anzahl an Koeffizienten zu haben. Hier hat sich gezeigt, dass das holographische Neuron eher für grosse Eingangsvektoren geeignet ist. Es war schwierig, ein gutes Verhältnis zwischen der Grösse des Eingangsvektors und der Grösse des Antwortvektors zu finden. Dabei ist zu beachten, dass eine Dimension des Antwortvektors kleiner als 20 vermutlich wenig Sinn macht, weil eine relativ grosse Anzahl an Phonemen unterschieden werden sollen. Die Abbildung 3.2 zeigt hier Lernschwierigkeiten mit dem CDF-Wavelet bei einer Tiefe von sieben. Wir vermuten, dass die gewählte Tiefe der Zerlegung nicht ausreichte.

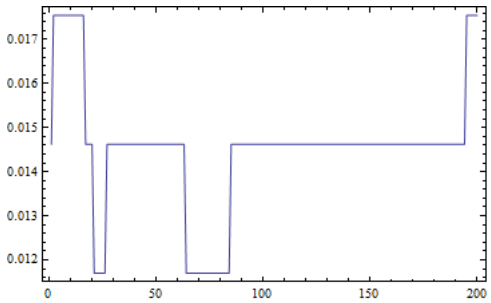

Abbildung 3.2: Performance-Index beim Lernen von Phonemen (Diskrete Wavelet-Koeffizienten).

3.1.3 Lernen der Kontinuierlichen Wavelet Koeffizienten

Je nach Phonembeispiel haben die Wavelet transformierten Daten verschieden viele Koeffizienten. Da sich die Unterschiede in Grenzen halten, können die x grössten Koeffizienten verwendet werden, wobei x die Anzahl Koeffizienten des kürzesten Phonems darstellt.
So konnte durch Aneinanderreihen der Wavelet Koeffizienten des in der Literatur oft verwendeten Morlet Wavelets mit sieben Oktaven und sieben Stimmen ein Eingangsvektor der Dimension 5000 gebildet werden.
Der von uns mit dem holographischen Neuron erzielte Lernerfolg hält sich jedoch in Grenzen. So konnte nach 16000 Epochen ein Performance Index von ca. 0.5 erreicht werden. Für diese 16000 Epochen brauchte unser Rechner mit 8GB Memory und einem Intel Xeon X3450 Prozessor allerdings rund zwei Wochen. Die 52.3 prozentige Erkennung des Phonems f sollte nicht zu euphorisch betrachtet werden. Sie resultiert vermutlich daraus, dass dieses Phonem in jeder Epoche als letztes gelernt wird.

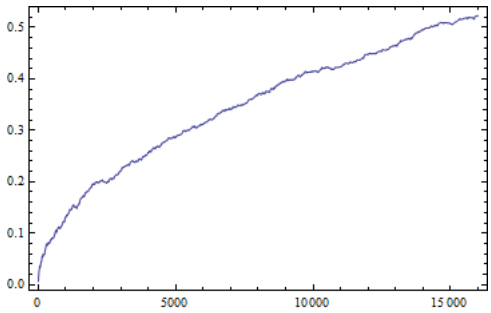

Abbildung 3.3: Performance-Index beim Lernen von Phonemen (Kontinuierliche Wavelet-Koeffizienten).

Phonem	**Erkennungsrate**
s	15.4 Prozent
sh	20.7 Prozent
z	20.4 Prozent
dh	16.3 Prozent
f	52.3 Prozent

Tabelle 3.1: Erkennungsrate nach 16000 Lernepochen und einem Performance-Index von 0.5.

3.2 Clustering

Unter Clustering versteht man das Verfahren, Datenobjekte mit ähnlichen Eigenschaften einer Gruppe zuzuordnen [11]. Damit kann man Eigenschaften erkennen, die uns helfen, die einzelnen Phonemgruppen zu unterscheiden. Das Verfahren scheint sinnvoll, da alle Samples eines Phonems einander in irgendeiner Weise ähneln. Wäre dies nicht der Fall, könnte das menschliche Gehör die verschiedenen Phoneme nicht auseinander halten. Wir haben darum zwei Clusteringverfahren getestet. Es zeigt sich, dass das k-Means Verfahren bei der Klassifizierung von Phonemen dem HLC unterlegen ist.

3.2.1 k-Means

Der k-Means Clustering Algorithmus ist ein schneller und leicht zu implementierender Algorithmus, welcher jeden Punkt dem ihm am nächsten liegenden Cluster zuordnet und danach die Cluster-Mittelpunkte neu berechnet. Dieses Verfahren zeichnet sich dadurch aus, dass seine Cluster-Schwerpunkte bei jedem Durchlauf wandern und sich so eindeutige Cluster bilden können [vgl. 11].

k-Means Algorithmus

1. Definiere Anzahl Durchläufe (Epochen)
2. n Cluster-Punkte zufällig wählen.
3. Jedes Sample wird dem Cluster zugeordnet, zu welchem es die geringste Distanz besitzt.
4. Für jeden Cluster wird der Schwerpunkt neu berechnet.
5. Wiederhole Schritte 3 und 4 für jede Epoche.

3.2.1.1 Klassifizierung anhand der Rohdaten

Am einfachsten ist es, die Rohdaten der .wav Datei zu clustern. Da die Samples aber eine unterschiedliche Länge besitzen, mussten diese erst normalisiert werden. Die Samples mit weniger als 2000 Datenpunkten werden mit Nullen erweitert, bzw. beschnitten, falls sie grösser als 2000 sind. Diese 2000 Datenpunkte entsprechen einer Dauer von 0.125 Sekunden.
Es wurden nur die fünf Phoneme s, sh, ch, dh und f betrachtet. Dem k-Means Algorithmus wurden fünf zufällig gewählte Cluster-Punkte vorgegeben und 200 Epochen wurden ausgeführt. Es konnten keine eindeutigen (reinen) Cluster ermittelt werden. Dies kann mehrere Gründe haben. Zum einen ist es möglich, dass nicht alle Phoneme aus der TIMIT Datenbank exakt getroffen wurden und diese dadurch zeitlich verschoben sein können. Auch kann durch das Beschneiden langer Phoneme charakteristische Information verloren gehen. Das Hauptproblem ist jedoch, dass sich die Phoneme der verschiedenen Sprecher in dieser Darstellung zu stark im Frequenzverhalten sowie in der Amplitude (Lautstärke) unterscheiden. Aus diesen Gründen waren auch die Cluster-Schwerpunkte zufällig und konnten sich nicht in einem Bereich einpendeln. Abschliessend kann man sagen, dass sich die wav-Darstellung nicht zum clustern mittels k-Means eignet.

3.2.1.2 Klassifizierung anhand der Kontinuierlichen Wavelet-Transformation

Wie im Kapitel 2.2 beschrieben, eignen sich Wavelets gut als Darstellungsform für Frikative, um deren Eigenschaften darzustellen. Wir waren überzeugt, die mit den Morlet-Koeffizienten dargestellten Phoneme gut clustern zu können. Die Ergebnisse waren jedoch ähnlich schlecht wie bei den Rohdaten. Auch durch variieren der Anzahl vorgegebener Cluster konnte keine Verbesserung der Resultate erzielt werden.

3.2.1.3 Klassifizierung anhand der Diskreten Wavelet-Transformation

Auch Clustern mit den Diskreten Wavelets ergab keine reinen Cluster. Verwendet wurde das CDF-Wavelet der Stufe 5 sowie das Daubechies-Wavelet der Stufe 5 und 6. Bei allen 3 Versuchen mit je 2000 Epochen waren die gefunden Cluster bunt durchmischt. Vermutlich sind die einzelnen Phonem-Cluster in einer oder mehreren Dimensionen verformt, was es für den k-Means Algorithmus praktisch verunmöglicht, diese zu trennen. Dies betrifft sowohl die kontinuierliche sowie die diskrete Darstellung.

3.2.2 Hebbian Learning Clustering

Der von der Stoop-Gruppe an der ETH Zürich entwickelte HLC Algorithmus basiert auf der Hebbschen Lernregel [5]. Er verstärkt die Verbindung zwischen zwei gleichzeitig feuernden Neuronen, bzw. baut diese ab, wenn sie nichts miteinander zu tun haben. Er funktioniert also ähnlich wie das menschliche Gehirn. Die Verwendung dieses Algorithmus brachte den entscheidenden Durchbruch.

Als Eingabe für den HLC-Algorithmus dienen die Koeffizienten des diskreten CDF-Wavelet. Das HLC Verfahren erstellt nun eine Distanzmatrix mit allen Elementen. Da dies ein enormer Speicherbedarf bei den über 2000 Samples ist, wurde aus Performancegründen die Anzahl Koeffizienten möglichst gering gehalten. Aus diesem Grund wurde für den HLC ein Wavelet mit der Tiefe 5 verwendet. Somit liefert die diskrete Wavelet-Transformation (nicht aber die diskrete Wavelet Packet Transformation) sechs starke Koeffizienten [12]. Der HLC-Algorithmus wurde bereits in Mathematica implementiert und uns zur Verfügung gestellt. Er musste jedoch geringfügig für unsere Problemstellung angepasst werden. Dies beinhaltete hauptsächlich die Parameteranpassungen, ersichtlich in der Tabelle 3.2.

Parameter	**Parameterwert**	**Beschreibung**
Neighbourhood Neurons	2	Anzahl gleichzeitig feuernde Neuronen
d0 Constant	3.4	Lokaler Parameter für den Algorithmus
tau Constant	2.6	Fenster für gleichzeitiges feuern von Neuronen
VoltageIR	2.2	Spannung
Theta	0.8	Grenzbereich
WaveletResize	3	Faktor um Waveletmerkmale besser Sichtbar zu machen

Tabelle 3.2: HLC Parameter.

Es kann so ein s-Cluster mit einem 90 prozentigen Reinheitsgrad gefunden werden. Nach ausklammern der bereits klassifizierten s-Samples können weitere Cluster gefunden werden, welche einen 70-90 prozentigen Reinheitsgrad besitzen. Hierbei ist anzumerken, dass Zischlaute wesentlich dominanter gegenüber nicht Zischlauten sind, was der Grund für das ausklammern ist.
Durch Bildung kleiner Phonem-Gruppen $\{s, z\}$, $\{sh, ch\}$ und $\{t, dh\}$ konnte eine durchschnittliche Erkennungsrate von 60 Prozent erreicht werden. Wie bereits erwähnt scheinen die Phonem-Cluster stark verformt zu sein. Deshalb wurden zur Distanzbestimmung drei Verfahren untersucht. Dazu gehören die Distanzbestimmung zum Cluster-Mittelpunkt mittels Euklidischer Norm, sowie der Abstand zum Cluster-Rand mittels der Euklidischen Norm und der Maximum Norm. Zur Zeit ist unklar, welches Distanzmass sich am besten eignet. Es wäre auch möglich, dass dies von Phonem zu Phonem variiert.
Die Verformung der Phonem-Cluster lässt darauf schliessen, dass es sich hierbei um so genannte mathematische Shrimps handelt [13]. Diese These wird auch durch den Erfolg des HLC-Verfahrens untermauert.

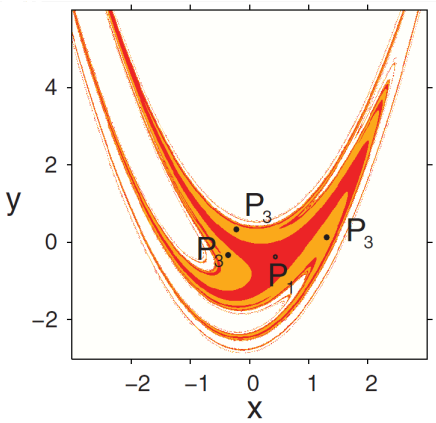

Abbildung 3.4: Beispiel eines Shrimps im $\mathbb{R}^2$.

Versuchsweise wurde die Tiefe des CDF-Wavelet auf 3 sowie 4 reduziert, um den Einfluss dieses Parameters zu überprüfen. Bei der Wavelet-Tiefe von 3 war es dem HLC nicht möglich, vernünftige Cluster zu finden. Mit dem Parameterwert 4 konnten bereits Cluster mit Phonem-Paaren gefunden werden. Diese Ergebnisse führen zur Annahme, dass eine höhere Wavelet-Tiefe bessere Resultate liefert. Somit kann der Wert 4 durchaus als Parameterschwellwert bezeichnet werden. Es ist zu beachten, dass die Grundparameter aus Tabelle 3.2 nicht angepasst wurden, was zusätzlich für die dramatisch verschlechterten Ergebnisse verantwortlich sein kann.

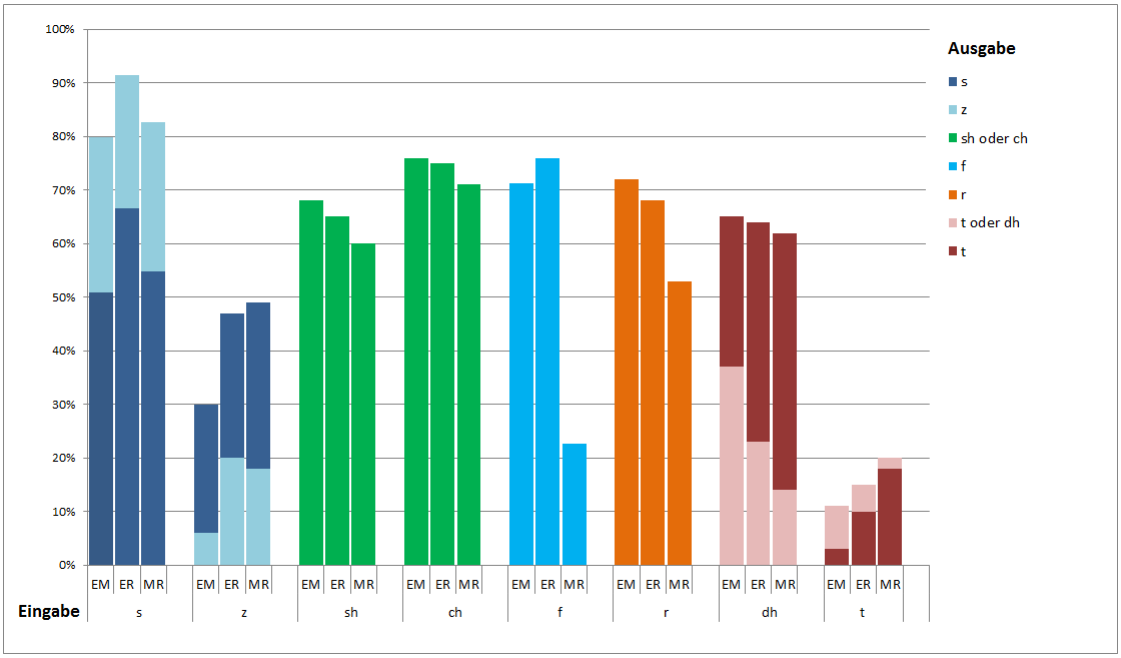

Abbildung 3.5: In dieser Grafik sieht man die verschiedenen Resultate pro Phonem und Distanznorm.

EM: Abstand zum Cluster-Mittelpunkt nach Euklidischer Norm.

ER: Abstand zum Cluster-Rand nach Euklidischer Norm.

MR: Abstand zum Cluster-Rand nach Maximum Norm.

Kapitel 4

Ausblick

Wir haben gezeigt, dass sich die Waveletdarstellung gut eignet, um Frikative zu unterscheiden. Weiter könnte abgeklärt werden, ob sich die Wavelets ebenso gut für andere Phonem-Gruppen (zB. Vokale) eignen.
Das holographische Neuron scheint eine Klassifizierung der Frikative durchaus zu ermöglichen. Abbildung 3.3 in Kapitel 3.1.3 zeigt, dass der Performance Index zwar mit den Lernepochen aufsteigt, jedoch flacht die Kurve nach oben immer weiter ab. Das bedeutet, dass der Performance Index nur durch einen grossen Mehraufwand verbessert werden kann. Hier müsste abgeklärt werden, ob ein Performance Index von Eins überhaupt erreicht werden kann. Dazu müsste man das Gesetz, dem das Anstiegsverhalten gehorcht, ermitteln.
Der HLC-Algorithmus ist in der Lage, die Phoneme zu unterscheiden. Da wir die ultimative Unterscheidung noch nicht erreicht haben, ist es sicher sinnvoll, noch mehr Zeit in das Clustering zu investieren. Die Darstellungsform könnte der Schlüssel sein um eine hervorragende Phonemerkennung zu ermöglichen. Als weiteren Schritt sollten die Phoneme noch mit anderen Wavelets als Basis geclustert werden. Auch sollten weiblichen Stimmen einbezogen werden, um zu sehen, wie gut das ganze System skaliert und wie gut die Wavelets die männliche und die weibliche Stimme normalisieren. Weiter sollte erprobt werden, welches die optimale Wavelet-Tiefe ist und gegebenenfalls die Parameter für den HLC mittels Gradientenabstieg bestimmen.

Kapitel 5

Persönliche Berichte

5.1 Rolf Schönenberger

Das Thema der Spracherkennung hat mich zunehmend gefesselt, da es für mich persönlich ein interessanter Lerneffekt bedeutete. Meine Hauptaufgabe war das Clustering, was mich fasziniert hat. Die frühen Erfolge des HLC waren für mich ein Motivationsschub, mich noch weiter in das Themengebiet einzuarbeiten. Es würde mich reizen, in die Spracherkennung noch mehr Zeit investieren zu können.

5.2 Claude Baumann

Mich hat die Spracherkennung im Allgemeinen sehr interessiert. Obschon es bereits seit mehreren Jahren diverse Ansätze gibt, so hat in der Öffentlichkeit die Sprachsteuerung bzw. Erkennung erst durch Apples SIRI an Bedeutung gewonnen.
Ich habe mich hauptsächlich in das Gebiet des holographischen Neurons eingearbeitet. Ich war erstaunt, wie viel Mühe unsere Rechner mit der Datenverarbeitung hatten. Gerne würde ich auch weiterhin neue Ansätze in diesem Bereich verfolgen.

Literaturverzeichnis

[1] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. 1989.

[2] D.B. Paul. Speech recognition using hidden markov models. *The Lincoln Laboratory Journal*, Volume 3, 1990.

[3] Besprechung vom 22.09.2011 mit Herrn Prof. Dr. R. Stoop.

[4] R. Stoop. *Wissensbasierte Systeme*, chapter Holographische Synapsen, pages 124–131. ETHZ, 2011.

[5] R. Stoop, F. Landis, and T. Ott. *Neural Computation*, (22):273–288, 2010.

[6] B.T. TAN, M. Fu, A. Spray, and P. Dermody. *The use of wavelet transforms in phoneme recognition*. The University of Newcastle, Australia.

[7] J.V. Bouvrie. *Hierarchical Learning: Theory with Applications in Speech and Vision*. Massachusetts Institute of Technology, 2009.

[8] C. Lopes and F. Perdigao. *Phone Recognition on the TIMIT Database*. Instituto de Telecomunicações, Portugal, 2011.

[9] Linguistic data consortium. URL `http://www.ldc.upenn.edu/`. zuletzt besucht am 15.12.2011.

[10] W. Bäni. *Wavelets: Eine Einführung für Ingenieure*. Oldenbourg, 2005. ISBN 3-486-57706-9.

[11] R. Stoop. *Wissensbasierte Systeme*, chapter Clustering-Gruppierverfahren, pages 155–176. ETHZ, 2011.

[12] Wolfram mathematica 8 documentation. URL `http://reference.wolfram.com/mathematica/`. zuletzt besucht am 15.12.2011.

[13] R. Stoop, P. Benner, and Y. Uwate. *Physical Review Letters*, (105), 2010.