

Erkennung und Klassifikation von akustischen Ereignissen zur häuslichen Pflege

Detection and Classification of Acoustic Events for In-Home Care

Dipl.-Phys. Jens Schröder, Dipl.-Ing. Stefan Wabnik, Dr. Peter W. J. van Hengel, Dipl.-Ing. Stefan Goetze, Fraunhofer Institut für Digitale Medientechnologie (IDMT), Projektgruppe Hör-, Sprach- und Audiotechnologie (HSA), Oldenburg, Germany, jens.schroeder@idmt.fraunhofer.de

Kurzfassung

Im Zuge des demographischen Wandels wird die Anzahl älterer Mitmenschen innerhalb der nächsten Jahrzehnte stetig ansteigen [1]. Im Bereich des „Ambient Assisted Living“ (AAL) werden unterstützende Systeme, Algorithmen und Methoden entwickelt, um älteren Menschen ein möglichst langes, selbstbestimmtes und sicheres Leben in ihren eigenen vier Wänden zu ermöglichen. Zu diesem Zweck können die Wohnungen mit verschiedenen Sensoren oder Geräten zum Sammeln und Evaluieren persönlicher Informationen ausgerüstet werden. Die akustische Überwachung durch Mikrophone wird durch die Nutzer im Vergleich zu einer Überwachung durch Kameras als geringerer Eingriff in die Privatsphäre empfunden. In diesem Beitrag wird ein akustisches Überwachungssystem zur Auswertung akustischer Signale exemplarisch anhand von vier Ereignissen (Husten, Klopfen, Klatschen und Telefonklingeln) vorgestellt. Die akustische Überwachung besteht aus mehreren, aufeinanderfolgenden Signalverarbeitungsschritten: Aus dem mittels Mikrophon aufgenommenen Signal wird das sogenannte *Cochleogramm* [2] berechnet, welches psycho-physiologisch motiviert ist. Aus dem Cochleogramm wird das Hintergrundrauschen geschätzt und in einem ersten Verarbeitungsschritt eliminiert. Nachfolgend werden diejenigen Abschnitte im Cochleogramm bestimmt, die ein einzelnes akustisches Ereignis beinhalten. Jedes dieser Ereignisse wird dann durch einen Klassifikationsalgorithmus ausgewertet. Während des Klassifikationsschritts wird die Zugehörigkeit zu einer Klasse überprüft, indem das Ereignis mit Repräsentanten verglichen wird, die zuvor aus einer *Self-Organizing Map* [3] ermittelt wurden. Es zeigt sich, dass bei einer Richtig-Positiv-Rate von über 79 % alle Geräusche bis auf Klopfen bei einer Falsch-Positiv-Rate von unter 4 % liegen.

Abstract

Due to the demographic change, the number of older people will grow steadily within the next decades [1]. Technical systems, methods and algorithms developed in the context of Ambient Assisted Living (AAL) aim to enable these people to live self-determined and safe in their own homes as long as possible. To ensure this, these homes may have to be equipped with e.g. a health monitoring system, special sensors or devices for gathering and evaluating individual information. In this paper, an acoustic monitoring system for evaluation of acoustic cues is presented and tested exemplarily on four acoustic events (cough, knock, clap and phone bell). The acoustic monitoring includes several consecutive signal processing steps: an audio signal picked up by a microphone is transformed in a psycho-physiologically weighted domain using a gammatone filterbank, resulting in a so called *cochleogram* [2]. From this cochleogram, background noise is estimated and eliminated. Further signal processing determines partitions of the cochleogram which are considered to form acoustic events. Each of these events is then evaluated with a classification algorithm. In this classification step, class membership is checked by comparing the event under inspection to representatives which have been determined during a training phase utilizing a self-organizing map [3]. It is shown that a true positive rate from over 79 % can be achieved where the false positive rate is smaller than 4 % except for the event knocking.

1 Einleitung

Der demographische Wandel in den Industrieländern führt zu einer stetig wachsenden Zahl älterer Menschen [1]. Modellberechnungen zeigen, dass bis 2030 die Hälfte der Bevölkerung in Deutschland älter als 47 Jahre sein wird [4]. Bereits jetzt sind ein Achtel der in Krankenhäusern behandelten Patienten, die einen Unfall im Haushalt hatten, über 65. Die Wahrscheinlichkeit eines solchen Haushaltsunfalls

nimmt mit dem Alter zu [4]. Trotzdem möchten mehr als 73 % der älteren Menschen in ihrem vertrauten Umfeld bleiben und so lange wie möglich in ihrer eigenen Wohnung leben [5]. Ein Umzug in ein Pflegeheim wird als Verlust der Eigenständigkeit und Verlust an Lebensqualität wahrgenommen [6]. Um ihnen ein sicheres Leben in ihrem gewohnten Umfeld zu ermöglichen, muss der Gesundheitszustand regelmäßig kontrolliert werden und eine Unterstützung bei Alltagsaufgaben muss gewährleistet sein.

Dies kann einen erheblichen Aufwand bedeuten und kostenintensiv sein, wenn diese Betreuung allein durch eine Pflegekraft erfolgen soll.

Eine automatische Überwachung (engl.: Monitoring) kann eine Unterstützung bei diesen Aufgaben bedeuten. Allerdings wird ein Überwachungsgerät wie eine Videokamera als Eindringen in die Privatsphäre wahrgenommen und deshalb sehr kritisch von den betroffenen Personen beurteilt.

Eine möglicherweise eher akzeptierte Alternative ist die akustische Überwachung durch Mikrophone und eine entsprechende Signalverarbeitung. Falls die Analyse des akustischen Signals automatisch in einem geschlossenen System geschieht, so dass keine Daten gespeichert werden müssen und niemand diese abhören kann, wird dies als geringerer Eingriff in die Privatsphäre empfunden. Ein weiterer Vorteil eines solchen akustischen Überwachungssystems wäre, dass Mikrophone ohne großen Aufwand ambient in eine Wohnung integriert werden können.

In den folgenden Abschnitten wird ein solcher Ereigniserkenner/-klassifikator basierend auf einem akustischen Signal und mit niedrigem Rechenaufwand vorgestellt.

In Abschnitt 2 wird der Algorithmus für den Erkenner und den Klassifikator beschrieben. Hier wird eine Unterscheidung zwischen der Erkennung eines Ereignisses, d. h. der Detektion eines Geräusches, das nicht zum Hintergrundrauschen gehört, und dessen Klassifikation gemacht, bei der das Ereignis einer definierten Klasse zugeordnet wird. Vor der Erkennung und Klassifikation wird eine psychophysiologisch motivierte Vorverarbeitung angewendet die zu einem Cochleogramm führt. Aus diesem wird das Hintergrundrauschen entfernt.

In Abschnitt 3 werden die Versuchsanordnungen und -umgebungen, in denen der Erkenner/Klassifikator getestet wurde, beschrieben. Vier verschiedene Geräusche sollen erkannt werden: Husten und Klopfen wurden exemplarisch getestet, da sie Indikatoren für den Gesundheitszustand darstellen können – Husten als Symptom von Atemwegserkrankungen und Klopfen als möglicher Hinweis auf Demenz.

Als weiteres Geräusch ist Händeklatschen getestet worden. Klatschen kann zur einfachen Steuerung und Interaktion mit technischen Geräten benutzt werden, wie es von älteren Menschen gewünscht wird [7,8]. In [9] ist diese Anwendung zum Aufwecken eines persönlichen Haushaltsassistenten, der mittels eines automatischen Spracherkennungssystems [10] gesteuert wird, integriert.

Als viertes Geräusch wurde Telefonklingeln getestet. Ein akustisch erkanntes Telefonklingeln könnte für Hörgeschädigte visualisiert werden um diese zu alarmieren. Mehr als 40 % älterer Menschen beklagen sich, dass sie Schwierigkeiten haben solche Alarmsignale von Haushaltsgeräten zu hören [11].

Die Ergebnisse dieser Experimente werden in Abschnitt 4 präsentiert.

Eine Zusammenfassung dieses Beitrages befindet sich in Abschnitt 5.

2 Erkenner und Klassifikator

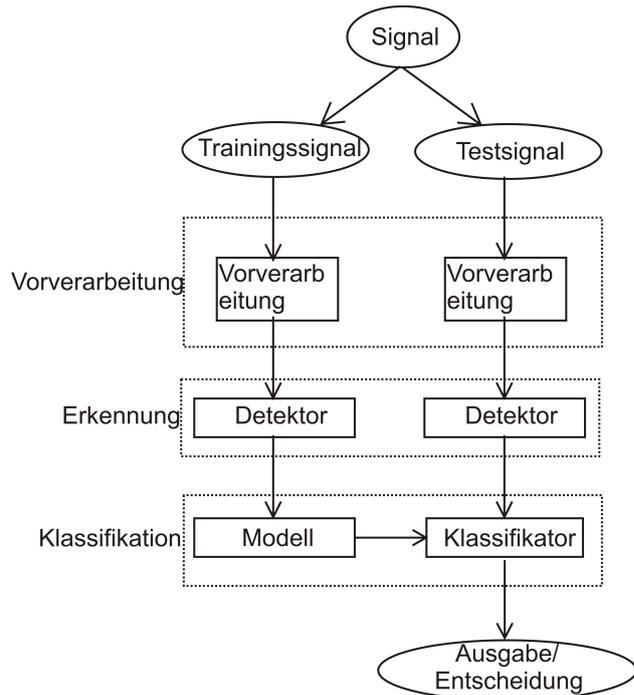


Abb. 1 Schema des Detektions- und Klassifikationsalgorithmus für die Trainings- und Testphase

Das akustische Erkenner-Klassifiziererpaar muss zu nächst trainiert werden bevor es benutzt werden kann. In der Trainingsphase wird der Erkenner/Klassifizierer auf richtig annotierten Beispieldaten trainiert, aus denen er Modellparameter lernt, die charakteristisch für die jeweilige Klasse sind.

In der Testphase werden andere, richtig annotierte Beispieldaten mittels der Klassenparameter, die in der Trainingsphase erlernt wurden, klassifiziert. In **Abb. 1** wird das Vorgehen schematisch veranschaulicht. In beiden Phasen wird ein Signal mittels eines Mikrophons aufgenommen. Dies aufgenommene Signal wird für den Erkenner und Klassifikator in eine geeignete Form vorverarbeitet. Im nächsten Schritt extrahiert der Erkenner Ereignisse aus dem kontinuierlichen Signal. Jedes dieser Ereignisse wird an den Klassifikator weitergeleitet. In der Trainingsphase werden Modelle aus den Ereignissen entwickelt. In der Testphase nutzt der Klassifikator die Modelle zur Entscheidung, zu welcher Klasse ein einzelnes Ereignis gehört.

2.1 Vorverarbeitung

Die Vorverarbeitung besteht hauptsächlich aus zwei Schritten: Eine Transformation des aufgenommenen Signals in ein Frequenz-Zeit-Muster, genannt Cochleogramm, und einer Vordergrund/Hintergrund-Trennung. Dies wird auf eine ähnliche Weise wie beim Aggressionsdetektor, beschrieben in [2], getan.

2.1.1 Cochleogramm als Frequenz-Zeit-Muster

Um ein akustisches Signal zu detektieren und klassifizieren, muss es erst durch Mikrophone aufgenommen und durch einen Analog-Digital-Wandler digitalisiert werden. Anschließend liegt es in einer quantisierten Amplituden-Zeit-Darstellung vor. Für unsere Zwecke ist dies aus verschiedenen Gründen nicht die beste Darstellung eines Geräusches. Das menschliche Ohr hat im Laufe der Zeit seine eigene Strategie zur Erkennung der Geräusche entwickelt, die für den Menschen wichtig sind. Im menschlichen Ohr wird das akustische Signal durch das Mittelohr verstärkt. Anschließend wird das Signal im Innenohr bzw. Hörschnecke (engl. cochlea) in sich überlappende Frequenzgruppen zerlegt. Diese Frequenzgruppen sind nicht linear über den Frequenzraum verteilt. Um dieses Verhalten zu simulieren, verwendet man eine Gammaton-Filterbank [12]. Für den hier vorgestellten Algorithmus wird eine Gammaton-Filterbank mit $M = 93$ Frequenzbändern implementiert. Die Mittenfrequenzen reichen von 20 Hz bis 8 kHz in 2.85 ERB Abständen verteilt um 1 kHz. Da die Phase des Signals in den meisten Fällen irrelevant ist, wird nach der Transformation in den logarithmischen Raum nur der Betrag weiter verarbeitet. Die Zeitauflösung des Filterbankausganges wird durch Unterabtastung reduziert, wodurch ein Ausgangsblock 5 ms darstellt.

Da das resultierende Frequenz-Zeit-Muster durch die cochleare Verarbeitung motiviert ist, wird es Cochleogramm genannt. In **Abb. 2** ist ein Cochleogramm einer Hustenserie aufgetragen.

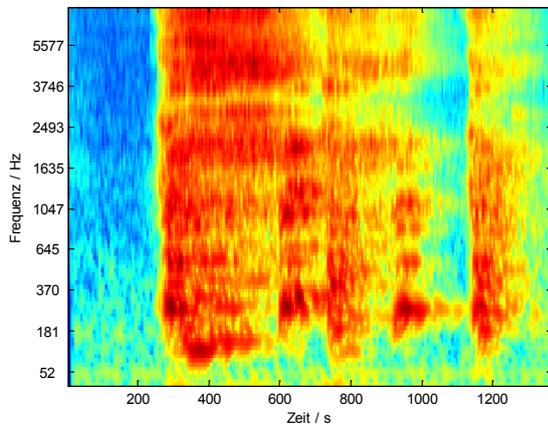


Abb. 2 Cochleogramm einer Hustenserie

2.1.2 Trennung von Vorder- und Hintergrund

Ein Cochleogramm $c(m,n)$, wobei m für die m -te Frequenzband von M und n für einen Zeitblock steht, kann als eine Zusammensetzung von Vordergrundgeräuschen, die zu detektieren sind, und Hintergrundrauschen betrachtet werden. In Freifeldumgebungen gibt es keine Kontrolle über die Art des Hintergrundrausens, so dass keine Annahmen über den Frequenzgehalt oder die Zeitentwicklung gemacht werden können. Die einzigen Annahmen, die hier gemacht werden, sind, dass das Hintergrundrauschen sich zeitlich weniger stark ändert als Vordergrundgerä-

usche und dass im Hintergrundrauschen keine scharfen Frequenzspitzen auftreten.

Um Vordergrundgeräusche vom Hintergrund zu trennen wird ein dynamisches Hintergrundmodell entwickelt, das den Hintergrundpegel zur Zeit n durch das Hintergrundmodell zur Zeit $n-1$ schätzt.

Zur Initialisierung des Modells werden die ersten I Zeitblöcke (25 Blöcke, 125 ms) gemittelt

$$bg(m,0) = \frac{1}{I} \sum_{i=1}^I c(m,i) \quad (1)$$

Für alle folgenden Zeitschritte wird das Hintergrundmodell zur Zeit $n-1$ als Schätzer für den nächsten Zeitblock n benutzt

$$p(m,n) = \sum_{i=1}^M \sigma_m(i) \cdot bg(i,n-1), \quad (2)$$

wobei $\sigma_m(i)$ ein Gewichtungsfaktor ist, der eine Verschmierung der Energien über die Frequenzbänder ermöglicht. Aus Normalisierungsgründen muss gelten:

$$\sum_{i=1}^M \sigma_m(i) = 1 \quad \forall m. \quad (3)$$

Normalerweise gilt

$$\sigma_m(i) = 0 \quad \forall i < m-2 \wedge i > m+2 \quad (4)$$

und

$$\sigma_m(m) \gg \sigma_m(i) \quad \forall i \neq m. \quad (5)$$

Beim weiteren Vorgehen wird ein Wahrscheinlichkeitsmaske erstellt, die den Unterschied zwischen vorhergesagtem Hintergrund $p(m,n)$ und dem Cochleogramm $c(m,n)$ gewichtet:

$$\mu(m,n) = 2^{-((c(m,n)-p(m,n))/\alpha(m))^6}, \quad (6)$$

wobei ein $\alpha(m)$ Parameter zur unterschiedlichen Gewichtung der Frequenzbänder ist. Dies ist notwendig, da die Gammaton-Filter unterschiedliche Bandbreiten umfassen. Die Maske adaptiert dynamisch das Hintergrundmodell:

$$bg(m,n) = (1-\beta) \cdot [\mu(m,n) \cdot c(m,n) + (1-\mu(m,n)) \cdot p(m,n)] + \beta \cdot p(m,n) \quad (7)$$

β stellt den Grad der Anpassung dar. Wird das Cochleogramm $c(m,n)$ gut durch den Schätzer $p(m,n)$ vorhergesagt, d.h. für die Differenz gilt

$$c(m,n) - p(m,n) \approx 0, \tag{8}$$

nimmt die Maske $\mu(m,n)$ Werte nahe eins an. In diesem Fall setzt sich das neue Hintergrundmodell aus dem Schätzer, der hauptsächlich aus dem alten Hintergrundmodell $bg(m,n-1)$ besteht, und einem gewissen Anteil vom aktuellen Cochleogramm $c(m,n)$ zusammen. Unterscheidet sich der Schätzer $p(m,n)$ hingegen deutlich vom aktuellen Cochleogramm $c(m,n)$, nimmt die Maske $\mu(m,n)$ Werte bei Null an. In diesem Fall besteht das neue Hintergrundmodell hauptsächlich aus dem alten Hintergrundmodell $bg(m,n)$. In **Abb. 3** ist das Hintergrundmodell der Hustenserie aus **Abb. 2** aufgetragen.

Die Vordergrundenergie ergibt sich zu

$$f(m,n) = (1 - \mu(m,n)) \cdot c(m,n). \tag{9}$$

Da die Hintergrundenergie dynamisch angepasst wurde, muss die Vordergrundenergie als diejenige Energie gewählt werden, die über der Hintergrundenergie liegt.

$$fg(m,n) = \begin{cases} f(m,n) & \forall f(m,n) > bg(m,n) \\ -\infty dB & \text{sonst} \end{cases} \tag{10}$$

Der separierte Vordergrund der Hustenserie aus **Abb. 2** ist in **Abb. 4** zu sehen.

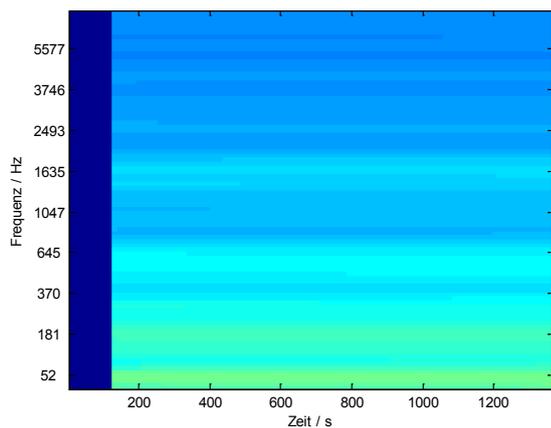


Abb. 3 Hintergrundmodell der Hustenserie aus **Abb. 2**

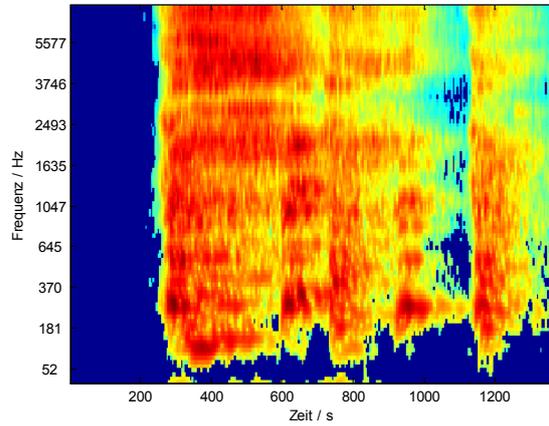


Abb. 4 Separierter Vordergrund der Hustenserie aus **Abb. 2**

2.1 Ereigniserkennung

Die Ereigniserkennung, d.h. der Erkenntnisgewinn, dass ein Geräusch vorliegt, das nicht zum Hintergrund gehört, wird durch einen Schwellwert abgeschätzt. Falls das Verhältnis zwischen Vorder- und Hintergrundenergie zu irgendeinem Zeitpunkt über einem definierten Schwellwert ist, wird dieser Zeitpunkt als Beginn des Ereignisses markiert. Das Ende wird markiert, wenn das Verhältnis wieder unter den Schwellwert fällt.

Um kurzzeitigen Energiefluktuationen um den Schwellwert entgegenzuwirken, wird eine zeitliche Hysterese eingeführt. Der Beginn eines Ereignisses wird dann erkannt, wenn das Energieverhältnis für eine längere Dauer als die Steigzeit größer als ein Schwellwert ist. Entsprechend gilt für das Auffinden des Endes, dass der Schwellwert für eine längere Dauer als die Fallzeit unterhalb des Schwellwertes liegen muss. Eine Skizze zur Veranschaulichung ist in **Abb. 5** zu sehen.

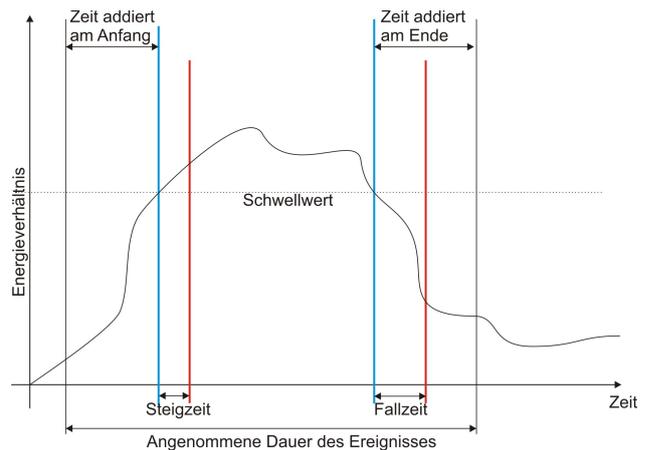


Abb. 5 Skizze des Entscheidungsschrittes des Ereigniserkenners. Bleibt das Energieverhältnis zwischen Vorder- und Hintergrund länger als die Steig- bzw. Fallzeit über/unter dem Schwellwert, wird ein Start- bzw. Endpunkt gesetzt.

2.2 Klassifikation

Es gibt zwei verschiedene Arbeitsphasen für den Klassifikator: Die Trainings- und die Testphase. Während der Trainingsphase wird aus den Trainingsdaten ein Modell entwickelt. In der Testphase wird das Modell zur Entscheidung, zu welcher Klasse ein Testdatum gehört, genutzt.

In beiden Phasen ist die Vorverarbeitung dieselbe. In diesem Artikel wird das separierte, Vordergrundcochleogramm verwendet, das mittelwertbefreit wurde, um möglichst pegelunabhängig zu sein.

Diese pegelunabhängigen Vordergrundcochleogramme wurden auf dieselbe Länge beschnitten, indem nur die ersten 180 ms ($N = 36$ Zeitblöcke) eines Ereignisses berücksichtigt wurden.

In der Trainingsphase wurde jede Klasse unabhängig von den anderen Klassen gelernt. Dazu ist ein Teil (ungefähr die Hälfte) der Trainingsdaten mit einer Self-Organizing Map (SOM) [3] gruppiert worden. Die Größe der SOM betrug 3×3 . Somit wurden die Trainingsdaten auf 9 für die Daten repräsentative Zentroide $fg_i(m, n)$ geschrumpft. Zentroide, die weniger als ein Datum repräsentierten, wurden vernachlässigt

Um ein Maß zur Abschätzung des Unterschiedes zwischen einem Testdatum und den Zentroiden zu haben, wurde die normalisierte L1-Norm gewählt

$$d_i = \frac{1}{M \cdot N} \sum_{m=1}^M \sum_{n=1}^N |fg(m, n) - fg_i(m, n)|. \quad (11)$$

Die minimale Distanz wurde als Maß der Klassenzugehörigkeit genommen.

Das bis jetzt noch nicht verwendete Trainingsmaterial wurde genutzt, um einen Schwellwert zu ermitteln, bis zu welchem eine minimale Distanz gerade noch klein genug ist für eine positive Klassifikation. Positive Klassifikation heißt, dass ein Testdatum zur getesteten Klasse klassifiziert wird. Als Schwellwert wurde das Minimum gewählt, das alle Trainingsdaten (mit Ausnahme von Ausreißern > 10 dB) positiv klassifizierte.

Während der Testphase muss dann nur noch die Differenz zwischen der minimalen Distanz und dem Schwellwert überprüft werden, um eine positive (Mitglied der getesteten Klasse) oder negative Klassifikation (nicht Mitglied der getesteten Klasse) zu erhalten.

In **Abb. 6** schematisiert die Klassifikationsaufgabe. Im Ereignisraum (grau) befinden sich neun Zentroide einer Klasse (hier vier eingezeichnet), die von ihren Schwellwertsphären (rot Kreise) umgeben sind. Falls sich ein Testdatum innerhalb einer dieser Sphären befindet, wird es positiv klassifiziert, ansonsten negativ.

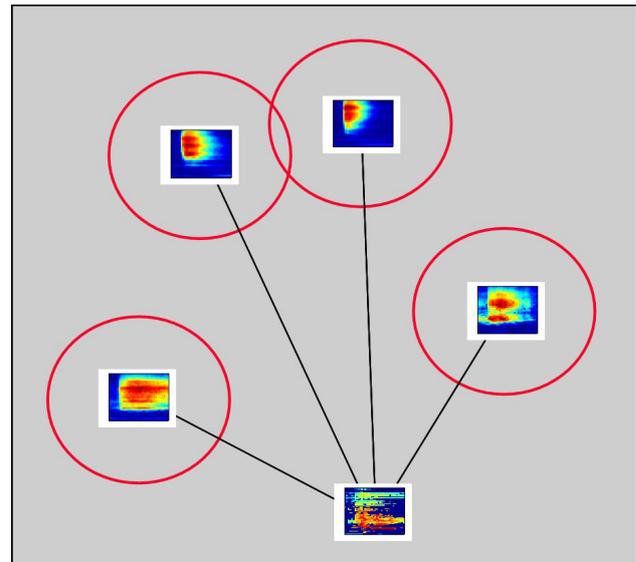


Abb. 6 Skizze des Klassifikationalgorithmus. Die neun Zentroide, die eine Klasse repräsentieren, befinden sich in einem hochdimensionalen Raum (grau). Sie werden von Sphären (rot Kreise) umgeben, die den Schwellwert darstellen. Befindet sich ein Testdatum innerhalb einer Sphäre, wird es positiv klassifiziert, ansonsten negativ.

3 Messaufbau

Zur Auswertung des Klassifikators wurden Aufnahmen in einer realen Umgebung gemacht. Im niedersächsischen Forschungsverbund „Gestaltung altersgerechter Lebenswelten“ [9] wurde eine komplette Wohnung eingerichtet um AAL-(Ambient Assisted Living-)Technologien zu testen. Zu diesem Zweck wurden in die Decke des Wohnzimmers Mikrophone installiert. Mit einem dieser Mikrophone wurden verschiedene Geräusche, erzeugt an unterschiedlichen Stellen im Raum, aufgenommen. Die Aufnahmen wurden in einer relativ ruhigen Umgebung durchgeführt. Die folgenden Geräusche wurden mit einer Abtastfrequenz von 48 kHz aufgenommen: Husten (zwei Personen), Händeklatschen (zwei Personen), Klopfen (an Türen, auf Tischen usw.) und Telefonklingeln (ein Telefon). Zusätzlich wurden noch einige Alltagsgeräusche aufgenommen (Konversation, Tütenrascheln usw.). Die Aufnahmen jeder Person und jeder Klasse wurden annotiert. Jede Aufnahmesitzung wurde mit dem Ereigniserkenner überprüft, der separierte Vordergrundcochleogramme für jedes Ereignis erzeugte. Der Erkenner hat 33 Husten von einer und 38 von der anderen Person gefunden, 48 bzw. 63 Klatschen der beiden Personen, 65 Klopfen, neun Telefonklingeln und 36 Alltagsgeräusche.

Dieses Material wurde in Trainings- und Testdaten eingeteilt um den Klassifikator zu testen. Für die Geräuschklassen Klatschen und Husten, die sehr personenspezifisch scheinen wurde die Daten einer Person zum Training und die der anderen zum Testen verwendet.

4 Evaluation

Es gibt verschiedene Möglichkeiten einen Klassifikator auszuwerten. In diesem Artikel werden die Richtig-Positiv-Rate und Falsch-Positiv-Rate berechnet.

Eine richtig-positiv Klassifikation bedeutet, dass das getestete Material und die Klasse, auf die getestet wird, übereinstimmen und der Klassifikator dieses richtig erkennt. Wird beispielsweise mit dem Klassifikator überprüft, ob ein Hustensignal als Husten erkannt wird und der Klassifikator entscheidet richtig, wird dies als „richtig-positiv“ gewertet. Wenn das Ereignis nicht mit dem Klassifikator übereinstimmt (Husten wird mit Klopfklassifikator überprüft) und das Ergebnis ist korrekter Weise negativ, wird die als „richtig-negativ“ bezeichnet. Die anderen beiden Klassifikationsergebnisse liefern falsche Ergebnisse. Ein Hustenereignis, dass mit dem Hustenklassifikator ausgewertet wird, aber nicht als Husten erkannt wird, bezeichnet man als falsch-negativ; ein Klopfen das mit dem Hustenklassifikator als Husten erkannt wird, wird falsch-positiv genannt.

In **Abb. 7 bis 10** sind die relative Anzahl der Ereignisse über die minimale Distanz aufgetragen. In **Abb. 7** wird auf Husten, in **8** auf Klopfen, in **9** auf Klatschen und in **10** auf Telefonklingeln klassifiziert. Die Daten, die zur getesteten Klasse gehören (im Folgenden Testklassendaten genannt), sind schwarz markiert, die restlichen grau.

Der geschätzte Klassifikationsschwellwert ist als vertikale Linie eingezeichnet. In den Legenden sind die Positiv-Raten angegeben.

Es ist ersichtlich, dass die Testklassendaten (schwarz) im Durchschnitt kleinere minimale Distanzen aufweisen als die anderen Daten. Eine Trennung der Gruppen ist möglich. Allerdings ist der geschätzte Klassifikationsschwellwert aus der Trainingsphase nicht optimal für die Trennung der Klassen.

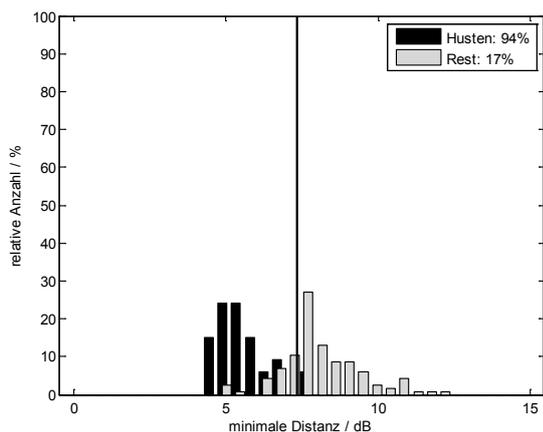


Abb. 7 Histogramm der minimalen Distanzen der Testklassendaten Husten (schwarz) und der anderen Daten (grau) zu den Hustenzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 7,3 dB an. In der Legende sind die Positivraten der Daten angegeben.

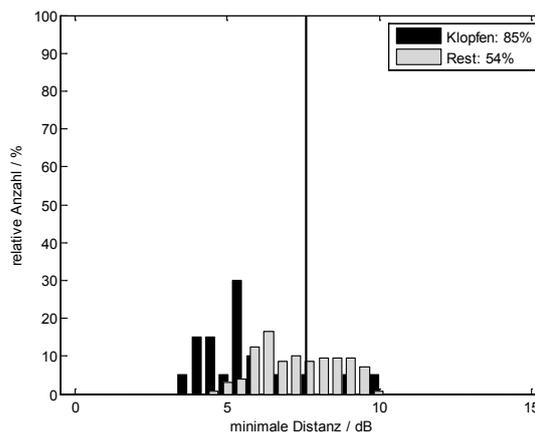


Abb. 8 Histogramm der minimalen Distanzen der Testklassendaten Klopfen (schwarz) und der anderen Daten (grau) zu den Klopfzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 7,6 dB an. In der Legende sind die Positivraten der Daten angegeben.

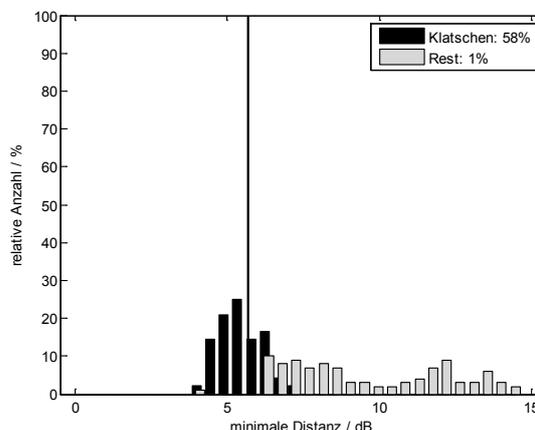


Abb. 9 Histogramm der minimalen Distanzen der Testklassendaten Klatschen (schwarz) und der anderen Daten (grau) zu den Klatschzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 5,7 dB an. In der Legende sind die Positivraten der Daten angegeben.

Beim Telefonklingeln ist der Klassifikationsschwellwert ziemlich gering (1,3 dB), obwohl der Abstand zwischen Testklassendaten und den anderen Daten groß genug ist, um einen höheren Schwellwert zuzulassen.

Ein fester Schwellwert bei 6,5 dB für alle Experimente liefert bessere Ergebnisse wie anhand von **Abb. 11 bis 14** zu sehen ist.

Die Richtig-Positiv-Rate ist bei allen Klassen über 79 %, die Falsch-Positiv-Rate lag bei allen Experimenten außer Klopfen unter 4 %. Klopfen scheint eine höhere Varianz innerhalb seiner Klasse zu haben, was an den verschiedenen Materialien und Klopfarten (flache Hand, Knöchel, Faust) liegen kann. Eine Repräsentation dieser Klasse durch neun Zentroide scheint zu wenig.

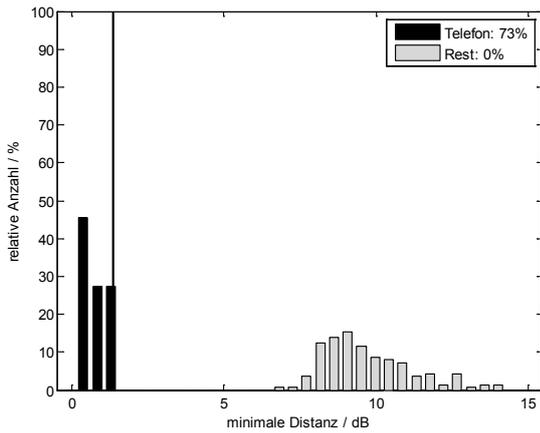


Abb. 10 Histogramm der minimalen Distanzen der Testklassendaten Klingeln (schwarz) und der anderen Daten (grau) zu den Klingelzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 1.3 dB an. In der Legende sind die Positivraten der Daten angegeben.

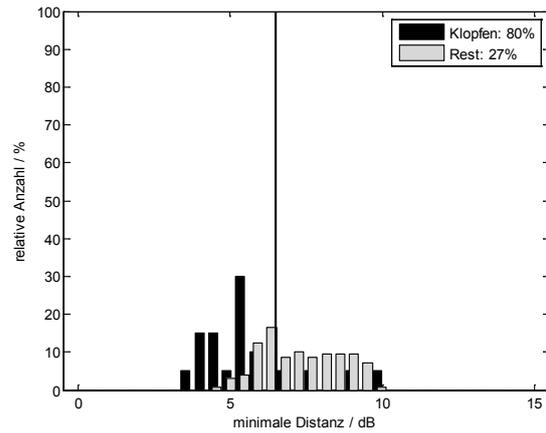


Abb. 12 Histogramm der minimalen Distanzen der Testklassendaten Klopfen (schwarz) und der anderen Daten (grau) zu den Klopfzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 6,5 dB an. In der Legende sind die Positivraten der Daten angegeben.

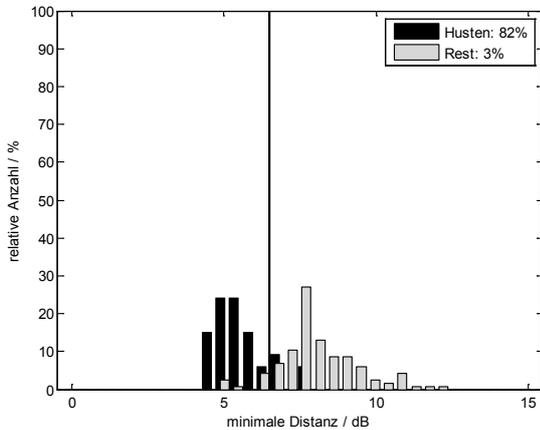


Abb. 11 Histogramm der minimalen Distanzen der Testklassendaten Husten (schwarz) und der anderen Daten (grau) zu den Hustenzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 6.5 dB an. In der Legende sind die Positivraten der Daten angegeben.

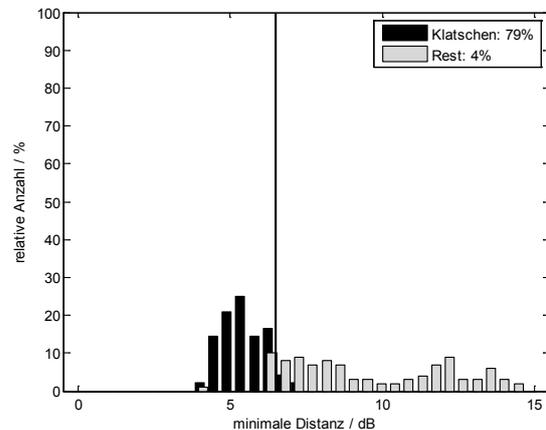


Abb. 13 Histogramm der minimalen Distanzen der Testklassendaten Klatschen (schwarz) und der anderen Daten (grau) zu den Klatschzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 6,5 dB an. In der Legende sind die Positivraten der Daten angegeben.

5 Zusammenfassung

In diesem Artikel wird ein Algorithmus zur Erkennung und Klassifikation akustischer Ereignisse unter realen Bedingungen vorgestellt. Die Vorverarbeitung findet unter psycho-physiologischen Gesichtspunkten statt. Der Vordergrund wird vom Hintergrundrauschen getrennt. Durch Pegelmessungen wird ein einzelnes Ereignis in einem kontinuierlichen Eingangssignal erkannt. Die Distanz zu durch eine SOM generierten Repräsentanten einer Klasse wird gemessen und bei Unterschreitung findet eine positive Klassifikation statt.

Der Erkenner/Klassifikator wurde in einer realen Umgebung mit Freifeldmikrofon in einer Decke getestet. Es zeigte sich, dass eine richtig-positiv Rate von über 79 % bei einer nur sehr geringen Falsch-Positiv-Rate von unter 4 % (mit Ausnahme von Klopfen) erreicht werden konnte.

6 Danksagung

Diese Arbeit wurde teilweise vom niedersächsischen Kultur- und Wissenschaftsministerium durch das Förderprogramm "Niedersächsisches Vorab" innerhalb des niedersächsischen Forschungsverbundes "Gestaltung altersgerechter Lebenswelten" gefördert.

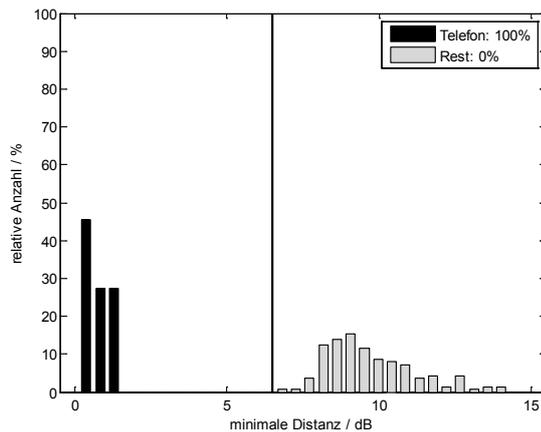


Abb. 14 Histogramm der minimalen Distanzen der Testklassendaten Klingeln (schwarz) und der anderen Daten (grau) zu den Klingelzentroiden. Die vertikale Linie zeigt den Klassifikationsschwellwert bei 6,5 dB an. In der Legende sind die Positivraten der Daten angegeben.

7 Literatur

- [1] European Commission Staff. Working Document. Europe's Demographic, Future: Facts and Figures. Commission of the European Communities; 2007 May
- [2] van Hengel, P. W. J. and Andringa, T. C.: Verbal aggression detection in complex social environments, AVSS '07: Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, 2007
- [3] Kohonen, T.: Self-Organizing Maps, 3rd Edition, Springer-Verlag Berlin Heidelberg, 2001
- [4] van Hengel, P. W. J. and Anemüller, J.: Audio Event Detection for In-Home Care, Int. Conf. on Acoustics (NAG/DAGA), 2009
- [5] Grauel J. and Spellberger A.: Akzeptanz neuer Wohntechniken für ein selbständiges Leben im Alter – Erklärung anhand soziostruktureller Merkmale, Technikkompetenz und Technikeinstellung, Zeitschrift für Sozialreform, Jg. 53, H. 2, pp. 191-215, 2007
- [6] Schneekloth U. and Wahl, H.-W.: Selbständigkeit und Hilfsbedarf bei älteren Menschen in Privathalten, Pflegearrangements, Demenz, Versorgungsangebote, Stuttgart: Kohlhammer, 2008
- [7] Alexandersson J, Zimmermann G, Bund J. User Interfaces for AAL: How Can I Satisfy All Users? In: Proc. Ambient Assisted Living - AAL. 2. Deutscher Kongress. Berlin: Germany; 2009. pp 5-9.
- [8] Helal S, Giraldo C, Kaddoura Y, Lee C, El Zabadani H, Mann W. Smart Phone Based Cognitive Assistant. In: Proceeding of The 2nd International Workshop on Ubiquitous Computing and Pervasive Healthcare Applications. Seattle; 2003. p 11.
- [9] Webseite des Forschungsprojekts Gestaltung altersgerechter Lebenswelten (GAL), <http://www.altersgerechte-lebenswelten.de>
- [10] Moritz, N.; Goetze, S.; Appell, J.-E.; Ambient Voice Control for a Personal Activity and Household Assistant, Proc. 4th German AAL-Congress, Berlin, Germany, Jan. 2011
- [11] Meis M, Appell J-E, Hohmann V, v Son N, Frowein H, Öster AM, Hein A. Tele-monitoring and Assistant System for People with Hearing Deficiencies: First Results from a User Requirement Study. Proceedings of European Conference on eHealth (ECEH); 2007. pp 163-175.
- [12] Hohmann, V.: Frequency Analysis and Synthesis using a Gammatone Filterbank, Acta Acustica United with Acustica, Vol. 88, pp. 433-442, 2002