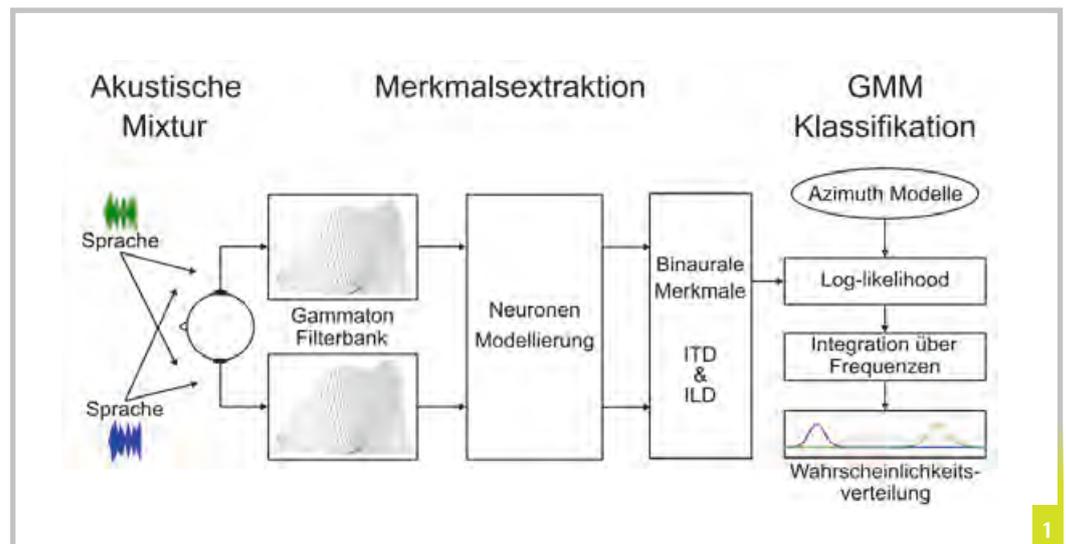


# Sprecherlokalisierung in Hörgeräten

Wie Hörgeräte Stimmen im Raum orten können

Die präzise räumliche Lokalisation von Sprechern im Umfeld eines Hörgerätes ist eine sehr wichtige und sehr rechenintensive Aufgabe, die die Qualität von digitalen Hörhilfen signifikant steigern kann.

In diesem Beitrag wird die enge interdisziplinäre Zusammenarbeit zwischen den Entwicklern der Lokalisations-Algorithmen an der Universität Oldenburg und den Hardware-Ingenieuren der Leibniz Universität Hannover präsentiert.



Weltweit leiden derzeit rund 466 Millionen Menschen an einem Hörverlust. Nach Schätzungen wird die Zahl bis 2050 auf über 900 Millionen Menschen steigen, also jeder Zehnte einen Hörverlust haben. Durch einen Hörverlust wird nicht nur das Sprachverstehen beeinträchtigt, sondern auch die akustische Orientierung im Raum anhand von Geräuschquellen oder Warnsignalen (z.B. im Straßenverkehr). Dies bedeutet einen Verlust an Lebensqualität, da es schwieriger ist, an alltäglichen Unterhaltungen teilzunehmen. In lauten Umgebungen mit vielen Sprechern – genannt Cocktail Party Problem – ist es für Normalhörende bereits schwierig, einer einzelnen Unterhaltung zu folgen. Für Hörgeschädigte ist es ohne Hilfsmittel, wie Hörgeräte, unmöglich einzel-

ne Gesprächspartner zu verstehen.

Die bisher verfügbaren Hörgeräte können nur in festgelegten Richtungen eine Hervorhebung eines Sprechers bewirken, also beispielsweise direkt frontal. Auch die Unterdrückung von Störquellen wie Verkehrslärm ist aufgrund der unterschiedlichen Klangcharakteristik möglich. Ein Vergleich aus dem Alltag macht die Problematik deutlich: Auf einem großen Parkplatz ein rotes Auto unter hunderten silbernen, schwarzen, blauen und grünen Autos zu finden ist einfach. Wenn auf dem Parkplatz aber ausschließlich rote Autos stehen und man möchte ein bestimmtes Modell eines Herstellers finden, wird es schwierig. Ähnlich geht es uns, wenn wir in einer

großen Menschenansammlung einen Sprecher von einem anderen akustisch unterscheiden wollen. Hier kommt die Sprecherlokalisierung ins Spiel, die einen Hinweis geben kann, in welcher Richtung man suchen soll.

Im Vergleich zu einer Brille wird das Tragen eines Hörgerätes gesellschaftlich noch immer als Behinderung angesehen. Daher wollen viele Betroffene kein Hörgerät tragen, wenn es nach außen sichtbar ist oder häufig die Batterien gewechselt werden müssen. Um diese Einschränkungen zu minimieren, müssen alle Komponenten, insbesondere der Prozessor für die Audiosignalverarbeitung, sehr klein und leistungsfähig sein sowie eine geringe Energieaufnahme besitzen.

Abbildung 1  
Blockdiagramm der Verarbeitungsschritte des verwendeten Lokalisationsalgorithmus

### Was bedeutet Lokalisation?

Der Begriff Lokalisation bedeutet die Ortung eines Objektes, also in welcher Richtung – vom Hörgeräteträger aus gesehen – sich ein Geräusch oder ein Sprecher befindet.

### Wie findet man ein Geräusch?

Um akustisch ein Geräusch zu lokalisieren, wird die Distanz zwischen linkem Ohr und der Geräuschquelle sowie zwischen dem rechten Ohr und der Geräuschquelle genutzt. Sobald sich die Geräuschquelle nicht genau frontal befindet, entsteht ein Zeitunterschied bei der Hörwahrnehmung, da der Schall mehr Zeit zum weiter entfernten Ohr benötigt. Zusätzlich zu diesem Unterschied der Ankunftszeit (engl. Interaural Time Difference – ITD) des Schalls hat auch der Kopf einen Einfluss. Gibt es keine direkte Sichtlinie zwischen der Schallquelle und einem Ohr, verringert sich die Lautstärke des Schalls (engl. Interaural Level Difference – ILD) für das durch den Kopf verdeckte Ohr.

Diese beiden Merkmale – also Lautstärke und Zeitunterschied des Schalls zwischen beiden Ohren – werden verwendet, um Geräusche zu lokalisieren. Für den Einsatz in Hörgeräten bedeutet das, dass diese Unterschiede entsprechend zwischen den Mikrofonen berechnet werden müssen.

### Der Algorithmus

Vom Department für Medizinische Physik und Akustik der Universität Oldenburg wurde aus einer Vorarbeit (Referenz T. May) ein Lokalisationsalgorithmus zur Verfügung gestellt, der für Hörgeräte entwickelt wurde und speziell für Umgebungen mit Wiederhall von Geräuschen geeignet ist.

Der Algorithmus besteht aus vier großen Stufen (vgl. *Abbildung 1*). Basierend auf der Frequenzauflösung des menschlichen Ohres werden die Audiosignale beider Mikrofone zunächst jeweils in 32 Frequenzbänder zerlegt. Diese Zerlegung wird mit einer sogenannten Gammaton Filterbank erzeugt. Das Innenohr mit den enthaltenen Haarzellen und den pulsartigen Neuronensignalen wird durch eine Einfachgleichrichtung und Kompression durch eine Wurzelfunktion modelliert. Anschließend wird eine Korrelation je Frequenzband berechnet, um den Zeitunter-

aktiven Schallquellen sind die Zuordnung von Sprach- und Störquellen, sowie die Bestimmung der Anzahl. Dies ermöglicht die Erkennung von bestimmten Umgebungen wie den öffentlichen Straßenverkehr, eine Wohnumgebung, den Arbeitsplatz, ein Konzert usw. sowie eine selbstständige Anpassung der Hörgeräte an diese Umgebungen.

### KAVUAKA – Der Forschungsprozessor für Hörgeräte

Für die Berechnung der Sprecherlokalisierung ist ein kleiner und energiesparender Prozes-

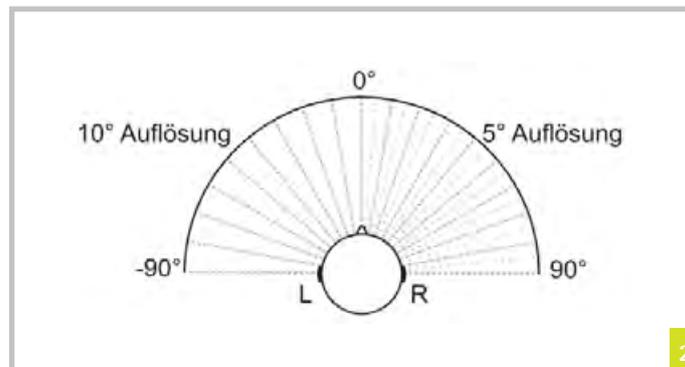


Abbildung 2  
Unterschiedliche Winkelauflösungen zur Lokalisation von Sprechern, die zur Optimierung genutzt werden.

schied zwischen den beiden Mikrofonensignalen zu bestimmen und parallel das Verhältnis der Lautstärken beider Signale (ILD) berechnet.

Danach wird ein zweidimensionales Gauss'sches Mixturen Modell (Gaussian Mixture Model - GMM) verwendet, um eine Wahrscheinlichkeitsverteilung aktiver Schallquellen über einen Winkel von 180° (frontal vom linken zum rechten Ohr – siehe *Abbildung 2*) mit einer Auflösung von 5° zu bestimmen. Das GMM wurde in einem Raum mit verschiedenen Konstellationen von Sprach- und Störquellen trainiert. Auf Basis der Wahrscheinlichkeitsverteilung ist es nun möglich, Algorithmen im Hörgerät richtungsabhängig zu steuern. Weitere Schritte nach der Lokalisierung von

sor nötig. Am Institut für Mikroelektronische Systeme (IMS) wurde der Algorithmus auf dem Forschungsprozessor KAVUAKA abgebildet. Durch Anpassungen in der Struktur des Prozessors ist es möglich, die rechenintensiven Teile des Algorithmus ausreichend schnell für schritthaltende Verarbeitung auszuführen. Bei der Implementierung ist der Speicherverbrauch des Gauss'schen Mixturen Modells eine wesentliche Herausforderung. Das vortrainierte Modell hat bei seiner Ausgangsgröße 5 2D-Gauss Parameter bei 15 Mixturenkomponenten, 37 Winkelklassen und 32 Frequenzbändern. Dies ergibt  $5 \cdot 15 \cdot 37 \cdot 32 = 88.800$  Parameter, die konstant im Speicher für die Erzeugung der Wahrscheinlichkeitsverteilung vorliegen müssen. Da diese mit

einer Genauigkeit von 32 Bit gespeichert werden, kann die Anzahl der notwendigen Speicherplätze in einem 64 Bit Speicher durch Kombination zweier 32 Bit Parameter in einem 64 Bit Speicherplatz halbiert werden. Es werden dann entsprechen 44.400 Speicherplätze benötigt. Das bedeutet, dass dieser Algorithmus in seiner ursprünglichen Version ohne Optimierungen nicht für den Hörgeräteprozessor

Symmetrien im zweidimensionalen Gauss'sches Mixturen Modell (GMM) identifiziert werden, welche auf der Anordnung der Mikrofone am Kopf basieren und sich dazu eignen, den Speicherbedarf des Modells zu reduzieren. Dazu gehört die links-rechts Symmetrie. Das bedeutet, dass für alle Winkel  $90^\circ$  von vorn bis zum linken Ohr ähnliche Parameter zu speichern sind, wie für die Winkel von vorn

quenzen weniger Mixturen benötigt werden, um eine ausreichende Genauigkeit zu erhalten. Durch eine logarithmisch verteilte Anzahl an Mixturen über die Frequenzen bleiben 16.330 Speicherplätze. In Kombination mit der Winkel- und Frequenzreduktion bleiben entsprechend 4.048 Speicherplätze. Es wurde also eine Reduktion des erforderlichen Speicherplatzes um etwa den Faktor 11 erzielt.

Abbildung 3  
Der Speicherbedarf des Lokalisationsalgorithmus wird in mehreren Schritten reduziert.



KAVUAKA mit 4096 64 Bit Speicherplätzen geeignet ist.

Für weitere Informationen zur Optimierung eines Prozessors für Hörgeräteanwendungen bietet der Artikel Chip Design für digitale Hörhilfen (Seitenverweis) eine detaillierte Beschreibung.

Eine wesentliche Herausforderung bestand (was typisch ist für die Wechselwirkung zwischen Algorithmen und Architekturen) darin, diesen komplexen und Speicherintensiven Sprecher-Lokalisations-Algorithmus so zu optimieren, dass er auf dem Hörgeräte-Prozessor implementiert werden kann.

### Reduktion des Speicherbedarfs

In Zusammenarbeit mit dem Department für Medizinische Physik und Akustik konnten

bis zum rechten Ohr. Somit muss nur eine Hälfte der Parameter wirklich gespeichert werden, während sich der Rest berechnen lässt. Ebenfalls berechnen lassen sich die  $5^\circ$  Winkelschritte auf Basis ihrer beiden Nachbarparameter. Somit ist es also ausreichend, die Parameter mit einer Winkelauflösung von  $10^\circ$  zu speichern. Zuletzt hat sich herausgestellt, dass auch die Parameter über die 32 Frequenzbänder eine lineare Abhängigkeit besitzen. Somit reicht es auch hier aus, nur die Hälfte der Parameter zu speichern und die fehlenden Parameter als Linearkombination zu berechnen. Es bleiben also nach diesen Optimierungen  $5 \cdot 15 \cdot 19 \cdot 16 = 22.800$  Parameter, entsprechend 11.400 Speicherplätze. Übrig bleibt, die 15 Mixturen je Winkel und Frequenz zu reduzieren. Dazu wird die Annahme verwendet, dass für niedrige Fre-

Alle Kombinationen und die dafür notwendige Anzahl an Speicherplätzen sind in *Abbildung 3* dargestellt. Nach diesen Reduktionen ist es möglich, auch bei geringem Speicherbedarf, aber gesteigertem Rechenbedarf, die Lokalisation durchzuführen.

Diese Arbeiten zeigen, dass in enger Wechselwirkung zwischen Algorithmikern und Hardware-Architekten Lösungen erarbeitet werden können, um komplexe leistungsfähige Algorithmen, die zu Beginn auf einer Hardware-Architektur nicht implementierbar sind, unter einem moderaten Qualitätsverlust dann doch zu realisieren.

Der berechneten Wahrscheinlichkeitsverteilung aktiver Schallquellen können nun weitere Analysen folgen. Dazu zählt die Unterscheidung von Sprach- und Stör-

quellen. Dadurch ist es möglich, die Umgebung zu klassifizieren – also zum Beispiel Verkehr, Zuhause, Konzert. Eine optimierte, effiziente Sprecherlokalisierung wird wesentlicher Bestandteil zukünftiger Hörgerätegenerationen sein.

### Literatur

- [1] T. May, S. van de Par, and A. Kohlrausch. "A Probabilistic Model for Robust Localization; Based on a Binaural Auditory Front-End". *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 1–13, 2011. ISSN 1558-7916. Doi: 10.1109/TASL.2010.2042128.
- [2] Seifert, C.; Thiemann, J.; Gerlach, L.; Volkmar, T.; Payá-Vayá, G.; Blume, H.; van de Par, S. (2017): Real-Time Implementation of a GMM-Based Binaural Localization Algorithm on a VLIW-SIMD Processor, *International Conference on Multimedia and Expo (ICME) 2017, IEEE*. Doi: 10.1109/ICME.2017.8019478
- [3] <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>



#### Prof. Dr.-Ing. Holger Blume

Jahrgang 1967, ist seit 2008 als Professor für „Architekturen und Systeme“ und als geschäftsführender Leiter des Instituts für Mikroelektronische Systeme (IMS) tätig. Seine wesentlichen Forschungsinteressen liegen auf dem Gebiet der Algorithmen und heterogenen Architekturen zur digitalen Signalverarbeitung, der Entwurfsraum-Exploration für diese Architekturen sowie den dazu erforderlichen Modellierungstechniken. Kontakt: [blume@ims.uni-hannover.de](mailto:blume@ims.uni-hannover.de)



#### Prof. Dr. Steven van de Par

Jahrgang 1966, ist seit 2010 Professor und Leiter der Abteilung Akustik am Department für Medizinische Physik und Akustik der Carl von Ossietzky University Oldenburg. Seine Forschungsinteressen liegen im Bereich der grundlegenden und angewandten Psychoakustik, dem Machine Listening und der technischen Akustik. Kontakt: [steven.van.de.par@uni-oldenburg.de](mailto:steven.van.de.par@uni-oldenburg.de)



#### Dr. Joachim Thiemann

Jahrgang 1973, ist seit 2018 angestellt beim European Research Center von Advanced Bionics in Hannover. Seine Forschungsinteressen liegen im Bereich von Machine Learning für Audio- und Biosignalverarbeitung, insbesondere Störgeräuschunterdrückung und Quellenlokalisierung. Kontakt: [joachim.thiemann@advancedbionics.com](mailto:joachim.thiemann@advancedbionics.com)



#### Christopher Seifert, M. Sc.

Jahrgang 1985, ist wissenschaftlicher Mitarbeiter und Doktorand am Institut für Mikroelektronische Systeme (IMS). Seine wesentlichen Forschungsinteressen sind die Optimierung von Prozessorarchitekturen und die effiziente Implementierung von Audiosignalverarbeitungs-algorithmen zur Sprecherlokalisierung im Bereich Hörgerätesysteme. Kontakt: [seifert@ims.uni-hannover.de](mailto:seifert@ims.uni-hannover.de)