(19) Deutsches Patent- und Markenamt

(10) **DE 10 2019 200 954 A1** 2020.07.30

(12) # Offenlegungsschrift

(21) Aktenzeichen: **10 2019 200 954.9**
(22) Anmeldetag: **25.01.2019**
(43) Offenlegungstag: **30.07.2020**

(51) Int Cl.: ***H04R 25/00*** (2006.01)
G10L 15/16 (2006.01)
G10L 17/18 (2013.01)
G10L 25/30 (2013.01)

(71) Anmelder:
Sonova AG, Stäfa, CH

(74) Vertreter:
RAU, SCHNECK & HÜBNER Patentanwälte Rechtsanwälte PartGmbB, 90402 Nürnberg, DE

(72) Erfinder:
Diehl, Peter Udo, Dr., 10585 Berlin, DE; Sprengel, Elias, 10625 Berlin, DE

(56) Ermittelter Stand der Technik:

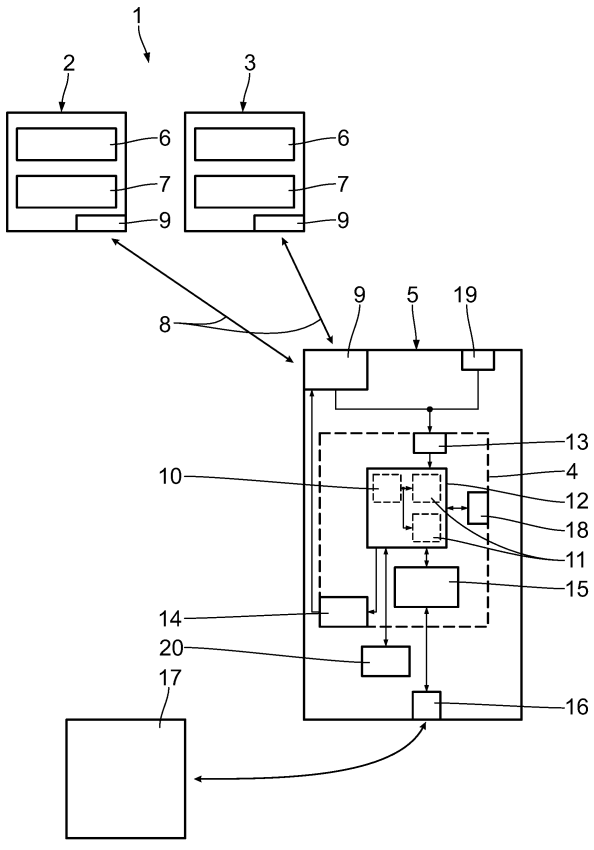
US	**2016 / 0 302 014**	**A1**
US	**2017 / 0 011 738**	**A1**
US	**2018 / 0 336 887**	**A1**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **Signalverarbeitungseinrichtung, System und Verfahren zur Verarbeitung von Audiosignalen**

(57) Zusammenfassung: Es wird eine Signalverarbeitungseinrichtung (4) zur Verarbeitung von Audiosignalen beschrieben. Die Signalverarbeitungseinrichtung (4) weist eine Eingangsschnittstelle (13) zum Empfangen eines Eingangssignals und eine Ausgangsschnittstelle (14) zum Ausgeben eines Ausgangssignals auf. Zudem weist die Signalverarbeitungseinrichtung (4) mindestens ein erstes neuronales Netzwerk (10) zum Aufbereiten des Eingangssignals und mindestens ein zweites neuronales Netzwerk (11) zum Separieren eines oder mehrerer Audiosignale aus dem Eingangssignal auf. Das mindestens eine erste neuronale Netzwerk (10) und das mindestens eine zweite neuronale Netzwerk (11) sind sequentiell angeordnet.



Beschreibung

[0001] Die Erfindung betrifft eine Signalverarbeitungseinrichtung zum Verarbeiten von Audiosignalen. Zudem betrifft die Erfindung ein System, insbesondere ein Hörgerätesystem, mit einer derartigen Signalverarbeitungseinrichtung. Die Erfindung umfasst zudem ein Verfahren zur Verarbeitung von Audiosignalen.

[0002] Signalverarbeitungsvorrichtungen und Verfahren zur Verarbeitung von Audiosignalen sind aus dem Stand der Technik bekannt. Sie finden beispielsweise Anwendung in Hörgeräten.

[0003] Es ist die Aufgabe der vorliegenden Erfindung, eine Signalverarbeitungseinrichtung zu schaffen, mit der eine Verarbeitung von Audiosignalen verbessert wird. Insbesondere soll eine Signalverarbeitungseinrichtung geschaffen werden, die eine effiziente Separierung eines Eingangssignals in einzelne oder mehrere Audiosignale erlaubt.

[0004] Diese Aufgabe ist gelöst durch eine Signalverarbeitungseinrichtung mit den in Anspruch 1 angegebenen Merkmalen. Die Signalverarbeitungseinrichtung weist eine Eingangsschnittstelle zum Empfangen eines Eingangssignals und eine Ausgangsschnittstelle zum Ausgeben eines Ausgangssignals auf. Zudem hat die Signalverarbeitungseinrichtung mindestens ein erstes neuronales Netzwerk zum Aufbereiten des Eingangssignals und mindestens ein zweites neuronales Netzwerk zum Separieren eines oder mehrerer Audiosignale aus dem Eingangssignal. Hier und im Folgenden ist der Begriff „neuronales Netzwerk“ als künstliches neuronales Netzwerk zu verstehen.

[0005] Der Kern der Erfindung besteht darin, dass das mindestens eine erste neuronale Netzwerk und das mindestens eine zweite neuronale Netzwerk sequentiell angeordnet sind. Die sequentielle Anordnung des mindestens einem ersten neuronalen Netzwerks und des mindestens einem zweiten neuronalen Netzwerks bedeutet, dass diese Netzwerke bei der Verarbeitung eines Eingangssignals hintereinander geschaltet sind. Insbesondere dient der Output des mindestens einen ersten neuronalen Netzwerks als ein Input für das mindestens eine zweite neuronale Netzwerk. Durch die sequentielle Anordnung der neuronalen Netzwerke ist eine funktionale Trennung in unterschiedliche Verarbeitungsschritte möglich. So erfolgt die Aufbereitung des Eingangssignals mit Hilfe des mindestens einen ersten neuronalen Netzwerks unabhängig von der Separation eines oder mehrerer der Audiosignale aus dem Eingangssignal mit Hilfe des mindestens einen zweiten neuronalen Netzwerks. Dies ermöglicht eine effiziente Verarbeitung der Audiosignale, insbesondere eine effiziente und genaue Separierung der Audiosignale. Die Separation der Audiosignale kann insbesondere in Echtzeit, dies bedeutet ohne nennenswerte Verzögerung, erfolgen. So kann beispielsweise die Aufbereitung des Eingangssignals durch das mindestens eine erste neuronale Netzwerk derart erfolgen, dass das aufbereitete Eingangssignal besonders einfach und effizient durch das mindestens eine zweite neuronale Netzwerk verarbeitet werden kann. Dies erhöht auch die Genauigkeit bei der Separation.

[0006] Ein weiterer Vorteil der erfindungsgemäßen Signalverarbeitungseinrichtung besteht in deren verbesserten Flexibilität. So können unterschiedliche erste neuronale Netzwerke mit verschiedenen zweiten neuronalen Netzwerken kombiniert werden, um eine an das jeweilige Eingangssignal angepasste Verarbeitung der Audiosignale zu gewährleisten. Als besonders effizient hat sich erwiesen, das mindestens eine erste neuronale Netzwerk unabhängig von dem Eingangssignal festzulegen, da die zur Aufbereitung des Eingangssignals nötigen Schritte universell für alle unterschiedlichen Arten von Eingangssignalen angewandt werden können. Das mindestens eine zweite neuronale Netzwerk kann dann besonders bevorzugt an die jeweiligen aus dem Eingangssignal zu separierenden Audiosignale angepasst werden.

[0007] Das Eingangssignal kann beispielsweise mit Hilfe einer oder mehrerer Aufzeichnungseinrichtungen aufgezeichnet werden und anschließend an die Eingangsschnittstelle der Signalverarbeitungseinrichtung übermittelt werden. Je Aufzeichnungseinrichtung weist das Eingangssignal beispielsweise einen oder mehrere Kanäle auf. Auf diese Weise können insbesondere Stereosignale aufgezeichnet werden.

[0008] Das Eingangssignal umfasst in der Regel eine unbekannte Anzahl unterschiedlicher Audiosignale. Die unterschiedlichen Audiosignale können insbesondere von unterschiedlichen Geräuschquellen, beispielsweise Gesprächspartnern, vorbeifahrenden Autos, Hintergrundmusik und/oder dergleichen stammen. Bevorzugt erfolgt die Separierung eines oder mehrerer Audiosignale aus dem Eingangssignal quellenspezifisch. In diesem Fall wird das Audiosignal einer bestimmten Geräuschquelle, beispielsweise eines Gesprächspartners, aus dem Eingangssignal separiert. Besonders bevorzugt werden mehrere Audiosignale aus dem Eingangssignal separiert. Auf diese Weise können die Audiosignale verschiedener Geräuschquellen unabhängig voneinander verarbeitet werden. Dies ermöglicht eine gezielte Verarbeitung und Gewichtung der einzelnen Audiosignale. Beispielsweise kann das Audiosignal eines Gesprächspartners verstärkt werden, während die Gespräche in der Nähe befindlicher Personen unterdrückt werden. Die Verarbeitung der Audiosignale ist quellenspezifisch möglich. Die Aufspaltung in einzelne Audiosignale, insbesondere in einzelnen

Geräuschquelle zugeordnete Audiosignale, mit Hilfe mindestens eines neuronalen Netzwerks ist unabhängig von der sequentiellen Anordnung von mindestens zwei unterschiedlichen neuronalen Netzwerken ein eigenständiger Aspekt der Erfindung.

[0009] Ein beispielhaftes Eingangssignal kann die letzten Millisekunden von kontinuierlich aufgezeichneten Audiodaten umfassen. Im Falle typischer Audiosignale mit 16000 Samples pro Sekunde kann das Eingangssignal beispielsweise etwa 128 Samples umfassen. Das Eingangssignal kann als Matrix dargestellt werden, deren Zeilenanzahl der Anzahl der Samples und deren Spaltenanzahl der Anzahl der Kanäle in dem Eingangssignal entsprechen.

[0010] Die Aufbereitung des Eingangssignals durch das mindestens eine erste neuronale Netzwerk kann als Teil eines Vorbereitungsschritts angesehen werden. Besonders bevorzugt erfolgt die Aufbereitung durch genau ein erstes neuronales Netzwerk. Dies hat sich als praktikabel erwiesen, da hierdurch eine einheitliche Handhabung des Eingangssignals, unabhängig von dessen Bestandteilen, beispielsweise den darin kombinierten Kanälen und/oder Audiosignalen, erfolgen kann. Zusätzlich zu der Aufbereitung mit Hilfe des mindestens einen ersten neuronalen Netzwerks kann eine klassische Aufbereitung des Eingangssignals erfolgen. Beispielsweise kann das Eingangssignal, insbesondere mehrere in dem Eingangssignal enthaltene Kanäle, normiert werden.

[0011] Die Aufbereitung des Eingangssignals hat den Vorteil, dass bei der Separierung von einem oder mehrerer Audiosignale aus dem Eingangssignal nicht mit einem Audioformat gearbeitet werden muss. Vielmehr ist es möglich, eine Repräsentation des Eingangssignals in Tensorform an das mindestens eine zweite neuronale Netzwerk zu übergeben. Hier kann eine effiziente und eindeutige Separierung erfolgen.

[0012] Das mindestens eine zweite neuronale Netzwerk kann eine variable Anzahl von Audiosignalen ausgeben. Bevorzugt hat das mindestens eine zweite neuronale Netzwerk eine feste Anzahl von Outputs. Im Fall mehrere zur Separation verwendeter zweiter neuronaler Netzwerke, kann jedes eine feste Anzahl von Outputs aufweisen. In diesem Fall gibt jedes zweite neuronale Netzwerk, das zur Separierung von Audiosignalen verwendet wird, eine feste Anzahl von aus dem Eingangssignal separierten Audiosignalen aus. Die Anzahl der separierten Audiosignale bemisst sich daher nach der Anzahl der zur Separierung verwendeten zweiten neuronalen Netzwerke und der jeweiligen Anzahl der Outputs. Beispielsweise können alle zweiten neuronalen Netzwerke drei Outputs aufweisen. Auf diese Weise können bei der Verwendung von beispielsweise zwei zweiten neuronalen Netzwerken zur Separierung bis zu sechs unterschiedliche Audiosignale aus dem Eingangssignal separiert werden. Es ist jedoch auch möglich, dass die unterschiedlichen zweiten neuronalen Netzwerke jeweils eine andere Anzahl von Outputs generieren. Auf diese Weise lässt sich die Anzahl der aus dem Eingangssignal mit Hilfe des mindestens einen zweiten neuronalen Netzwerks separierten Audiosignale noch flexibler festlegen.

[0013] Die von den zweiten neuronalen Netzwerken ausgegebenen Audiosignale können beliebig codiert sein. Ein weiterer Vorteil der sequentiellen Ausführung von Aufbereitung und Separierung ist jedoch, dass die Outputs des mindestens einen zweiten neuronalen Netzwerks selbst als Audiodaten oder eine Vorstufe von Audiodaten codiert sein können. Es ist möglich, dass das Audiosignal selbst durch das mindestens eine erste neuronale Netzwerk zur Verwendung in mindestens einem zweiten neuronalen Netzwerk optimiert oder aufbereitet wird. Beispielsweise kann ein aufbereitetes Audiosignal, das das erste neuronale Netzwerk ausgibt, durch das mindestens eine zweite neuronale Netzwerk in eine Vielzahl von neuen Audiosignalen umgewandelt werden. Dies bedeutet, dass das mindestens eine zweite neuronale Netzwerk generativ arbeiten kann.

[0014] Gemäß einem vorteilhaften Aspekt der Erfindung weist die Signalverarbeitungseinrichtung eine Mehrzahl zweiter neuronaler Netzwerke auf, wobei jedes der zweiten neuronalen Netzwerke an eine bestimmte Art von Audiosignalen angepasst ist. Dies ermöglicht eine besonders effiziente Separation von bestimmten Arten von Audiosignalen aus dem Eingangssignal. Durch die Mehrzahl zweiter neuronaler Netzwerke, die an unterschiedliche Arten von Audiosignalen angepasst sind, ist die Signalverarbeitungseinrichtung besonders flexibel und universal einsetzbar. Die Separation der Audiosignale kann durch einzelne oder mehrere der Mehrzahl zweiter neuronaler Netzwerke erfolgen. Das zur Separation verwendete zweite neuronale Netzwerk kann je nach Eingangssignal oder sonstigen Anforderungen aus der Mehrzahl der zweiten neuronalen Netzwerke auswählbar sein.

[0015] Die unterschiedlichen Arten von Audiosignalen bestimmen sich beispielsweise anhand deren jeweiligen Geräuschquellen, beispielsweise menschliche Sprecher oder Kraftfahrzeuge. Die Art der Geräuschquellen kann auch durch eine bestimmte Umgebung, beispielsweise Straßen- und Verkehrslärm oder Hintergrundmusik in einem Einkaufszentrum, bestimmt sein. Die Anpassung der zweiten neuronalen Netzwerke an die jeweilige Art von Audiosignalen erfolgt durch Trainieren der neuronalen Netzwerke, beispielsweise anhand von Datensätzen, die derartige Audiosignale enthalten.

[0016] Gemäß einem vorteilhaften Aspekt der Erfindung werden mindestens zwei, drei, vier oder mehr zweite neuronale Netzwerke parallel zur Separie-

rung von Audiosignalen aus dem Eingangssignal verwendet. Dies ermöglicht es, eine große Anzahl unterschiedlicher Audiosignale aus dem Eingangssignal zu separieren. Zudem ist die Flexibilität erhöht, da auf unterschiedliche Arten von Audiosignalen spezialisierte zweite neuronale Netzwerke kombiniert werden können, sodass die Separierung für unterschiedliche Arten von Audiosignalen auf einfache und eindeutige Weise erfolgen kann. Bevorzugt dient der Output des mindestens einen ersten neuronalen Netzwerks als Input, insbesondere als identischer Input, für alle zur Separierung parallel verwendeten zweiten neuronalen Netzwerke. Hierdurch ist sichergestellt, dass die verschiedenen Audiosignale zuverlässig aus dem Eingangssignal separiert werden.

[0017] Gemäß einem weiteren vorteilhaften Aspekt der Erfindung ist das mindestens eine zweite neuronale Netzwerk austauschbar. Das zur Separierung der Audiosignale verwendete mindestens eine zweite neuronale Netzwerk ist insbesondere auswählbar aus einer Mehrzahl zweiter neuronaler Netzwerke, die auf unterschiedliche Arten von Audiosignalen spezialisiert sind. Durch die Austauschbarkeit des mindestens einen zweiten neuronalen Netzwerks kann die Signalverarbeitungseinrichtung flexibel an die jeweiligen Eingangssignale angepasst werden. Durch die Wahl des jeweils geeigneten mindestens einen zweiten neuronalen Netzwerks ist auch die Genauigkeit beim Separieren der Audiosignale aus dem Eingangssignal verbessert. Weiterhin können die mehreren zweiten neuronalen Netzwerke, insbesondere auf einem AI-Chip, parallel ausgeführt werden. Die Bearbeitungszeit des Signals ist weiter reduziert.

[0018] Bevorzugt sind einzelne oder mehrere der Mehrzahl der zweiten neuronalen Netzwerke unabhängig voneinander austauschbar.

[0019] Die sequentielle Anordnung des mindestens einen ersten neuronalen Netzwerks und des mindestens einen zweiten neuronalen Netzwerks hat insbesondere beim Austausch des zweiten neuronalen Netzwerks denn Vorteil einer verbesserten Konsistenz des Signals. Beispielsweise werden Informationen im mindestens einen ersten neuronalen Netzwerk gespeichert und gehen beim Austausch nicht verloren. Eine Unterbrechung der Audiosignale und deren Separierung ist vermieden.

[0020] Gemäß einem weiteren vorteilhaften Aspekt der Erfindung sind das mindestens eine erste neuronale Netzwerk und das mindestens eine zweite neuronale Netzwerk Teil eines gemeinsamen neuronalen Netzwerks. Eine derartige Signalverarbeitungseinrichtung ist besonders effizient. Beispielsweise können das mindestens eine erste neuronale Netzwerk und das mindestens eine zweite neuronale Netzwerk gemeinsam ausgeführt werden, insbesondere auf einem einzelnen Prozessor, insbesondere auf einem AI-Chip. Das mindestens eine erste neuronale Netzwerk kann als Körper des gemeinsamen neuronalen Netzwerks gesehen werden, während das mindestens eine zweite neuronale Netzwerk als austauschbarer Kopf des gemeinsamen neuronalen Netzwerks fungiert. Das gemeinsame neuronale Netzwerk kann insbesondere eine Mehrzahl von zweiten neuronalen Netzwerken aufweisen, die flexibel und unabhängig voneinander austauschbar sind. Das gemeinsame neuronale Netzwerk wird in diesem Fall auch als ein neuronales Netzwerk mit rotierenden Köpfen bezeichnet.

[0021] Das Zusammenfassen des mindestens einen ersten neuronalen Netzwerks und des mindestens einen zweiten neuronalen Netzwerks in einem gemeinsamen neuronalen Netzwerk hat weiterhin den Vorteil, dass der Output des mindestens einem ersten neuronalen Netzwerks direkt als Input an das mindestens eine zweite neuronale Netzwerk übergeben wird. Eine zusätzliche Ausgabe und/oder Umwandlung des Outputs des mindestens einen ersten neuronalen Netzwerks ist vermieden.

[0022] Das mindestens eine erste neuronale Netzwerk und das mindestens eine zweite neuronale Netzwerk werden bevorzugt zunächst gemeinsam trainiert. Nachdem das mindestens eine erste neuronale Netzwerk auf die Aufbereitung des Eingangssignals ausreichend spezialisiert ist, genügt es, das mindestens eine zweite neuronale Netzwerk weiter auf die Separierung bestimmter Arten von Audiosignalen hin zu trainieren. Das mindestens eine erste neuronale Netzwerk kann in dieser Trainingsphase unverändert belassen werden.

[0023] Für das Trainieren unterschiedlicher zweiter neuronaler Netzwerke können verschiedene Datensätze verwendet werden. Beispielsweise wird eines der zweiten neuronalen Netzwerke auf die Separierung von weiblichen Stimmen und ein anderes zweites neuronales Netzwerk auf die Separierung von Warnsignalen im Straßenverkehr spezialisiert. Die zweiten neuronalen Netzwerke geben jeweils Audiosignale der Art aus, auf welche sie spezialisiert sind. Ein auf weibliche Stimmen trainiertes zweites neuronales Netzwerk wird daher eine weibliche Stimme erkennen und ein entsprechendes Audiosignal ausgeben. Jedes zweite neuronale Netzwerk weist bevorzugt eine Mehrzahl von Outputs auf. Ein auf weibliche Stimmen trainiertes zweites neuronales Netzwerk mit mehreren Outputs, kann mehrere Audiosignale, die verschiedenen weiblichen Stimmen entsprechen ausgeben. Weist ein zweites neuronales Netzwerk mehr Outputs auf als das Eingangssignal Audiosignale der Art hat, auf welche dieses zweite neuronale Netzwerk spezialisiert ist, können weitere Outputs des zweiten neuronalen Netzwerks auch an-

dere Arten von Audiosignalen enthalten, auf welche das zweite neuronale Netzwerk nicht trainiert ist. Andererseits können die zweiten neuronalen Netzwerke auch so trainiert sein, dass sie nur Audiosignale der Art, auf welche sie spezialisiert sind, ausgeben. Beispielsweise würde ein auf weibliche Stimmen spezialisiertes Netzwerk keine männlichen Stimmen ausgeben. Ist die Anzahl der Outputs eines zweiten neuronalen Netzwerks höher als die Anzahl der Audiosignale der Art, auf die das zweite neuronale Netzwerk trainiert ist, können die überschüssigen Outputs ein leeres Signal ausgeben. Das leere Signal entspricht einem Audiosignal, das keine Geräusche, also lediglich Stille, enthält. Enthalten viele Outputs ein derartiges leeres Signal, kann die Anzahl der verwendeten zweiten neuronalen Netzwerke reduziert werden. Das Verfahren ist effizient und stromsparend. Dies ist besonders vorteilhaft für mobile Anwendungen.

[0024] Alternativ können die zweiten neuronalen Netzwerke dazu trainiert werden, mögliche weitere Audiosignale in einem Restsignal gebündelt auszugeben. Beispielsweise kann ein auf weibliche Stimmen spezialisiertes Netzwerk männliche Stimmen, Straßenlärm und weitere Audiosignale zusammen als zusätzliches Restsignal ausgeben. Das Restsignal kann als Maß für nicht separierte Audiosignale dienen. Umfasst ein derartiges Restsignal noch eine große Anzahl von Informationen, kann die Anzahl der zweiten neuronalen Netzwerke und/oder die Anzahl von Outputs je zweitem neuronalen Netzwerk erhöht werden. Hierdurch kann die Anzahl der separierten Audiosignale einfach und flexibel an das Eingangssignal, insbesondere die Anzahl darin enthaltener Audiosignale angepasst werden.

[0025] Verschiedene zweite neuronale Netzwerke können durch das Training auch zu einem unterschiedlichen Grad spezialisiert werden. Beispielsweise ist möglich, ein zweites neuronales Netzwerk auf Stimmen allgemein zu trainieren und weitere zweite neuronale Netzwerke jeweils nur auf eine bestimmte Art von Stimme (tief, hoch, deutsch, englisch, etc.) zu trainieren. In diesem Fall kann das zweite neuronale Netzwerk, das Stimmen allgemein erkennt, verwendet werden, solange nur wenige Stimmen detektiert werden. Steigt die Anzahl der detektierten Stimmen können mehrere der stärker spezialisierten zweiten neuronalen Netzwerke eingesetzt werden. Die Anzahl der separierten Audiosignale kann flexibel angepasst werden.

[0026] Für das mindestens eine erste neuronale Netzwerk und das mindestens eine zweite neuronale Netzwerk können unterschiedliche Netzwerkarchitekturen verwendet werden. Die verwendete Architektur der neuronalen Netzwerke ist für die Separation der Audiosignale aus dem Eingangssignal nicht wesentlich. Als besonders geeignet haben sich jedoch Long Short-Term Memory (LSTM) Netzwerke erwiesen. Eine derartige Architektur ist besonders vorteilhaft, wenn das Eingangssignal jeweils nur wenige Millisekunden von längeren, insbesondere kontinuierlich aufgezeichneten Audiodaten ist. Eine LSTM Architektur des mindestens einen ersten neuronalen Netzwerks erlaubt es, Informationen über eine längere Zeitspanne der Audiodaten über längere Zeit zu speichern. Die gespeicherten Informationen können dann auch einem zuvor nicht verwendeten zweiten neuronalen Netzwerk übergeben und dort weiterverarbeitet werden. Hierdurch ist insbesondere möglich das mindestens eine zweite neuronale Netzwerk optimal zu initialisieren.

[0027] In einer bevorzugten Architektur kann das mindestens eine erste neuronale Netzwerk eine 1D Konvolutionsebene, auch 1D convolutional Layer genannt, und mindestens eine LSTM Ebene aufweisen. Besonders bevorzugt weist das mindestens eine erste neuronale Netzwerk eine 1D Konvolutionsebene und zwei LSTM Ebenen mit beispielsweise 1024 bzw. 512 Einheiten auf. Das Eingangssignal kann so in eine neue, kompaktere Repräsentation gebracht werden. Zwischen verschiedenen Ebenen können zudem sogenannte Skip Connections vorhanden sein. Dies erlaubt den Zugriff auf das originale Eingangssignal und auf alle Zwischenresultate. Zur Aufbereitung kann das Eingangssignal mittels der 1D Konvolutionsebene umgewandelt und mittels einer oder mehrerer LSTM Ebenen verbessert werden.

[0028] In einer bevorzugten Architektur kann das mindestens eine zweite neuronale Netzwerk mindestens eine LSTM Ebene und mindestens eine vollständig verknüpfte Ebenen, auch Dense Layer genannt, aufweisen. Ein beispielhaftes zweites neuronales Netzwerk kann beispielsweise zwei LSTM Ebenen mit 265 bzw. 128 Einheiten und zwei sich hieran anschließende vollständig verknüpften Ebenen mit 128 bzw. 64 Einheiten aufweisen. An die vollständig verknüpften Ebenen kann sich eine Konvolutionsebene, auch convolutional Layer genannt, anschließen. Eine derartige Architektur erlaubt die Ausführung des mindestens einen zweiten neuronalen Netzwerks mit gängiger Hardware. Beispielsweise benötigen ein erstes neuronales Netzwerk und drei zweite neuronale Netzwerke mit den jeweils oben beschriebenen bevorzugten Architekturen zur Ausführung eine Rechenleistung von 0,6 Terraflops. Gängige Mobilfunkgeräte weisen AI-Chips mit 2 oder mehr Terraflops, beispielsweise 5 Terraflops, auf.

[0029] Bei einer Mehrzahl von zweiten neuronalen Netzwerken können diese die gleiche oder unterschiedliche Architekturen aufweisen. In verschiedenen zweiten neuronalen Netzwerken kann die Anzahl der Ebenen und Einheiten variieren. Insbesondere die Anzahl der Einheiten kann abhängig von der Spezialisierung des jeweiligen neuronalen Netzwerks sein. Beispielsweise kann über ein Training

mit spezialisierten Datensätzen eine Reduktion der Einheiten erfolgen. Ein zweites neuronales Netzwerk, das beispielsweise nur auf Straßenlärm trainiert ist, kann eine wesentlich kleinere Architektur aufweisen, als ein zweites neuronales Netzwerk, das auf allgemeine Störgeräusche trainiert ist. Insbesondere bei der Verwendung mehrerer zweiter neuronaler Netzwerke kann deren Architektur vorteilhafterweise klein sein. Dies erhöht die Effizienz der zweiten neuronalen Netzwerke. Die neuronalen Netzwerke der Signalverarbeitungseinrichtung können beispielsweise auf beliebigen Prozessoren ausgeführt werden. Spezialisierte AI-Chips sind nicht zwingend erforderlich.

[0030] Gemäß einem weiteren vorteilhaften Aspekt der Erfindung weist die Signalverarbeitungseinrichtung eine Nutzerschnittstelle zum Empfang von Nutzereingaben und/oder zur Ausgabe von Informationen an einen Nutzer auf. Durch die Nutzerschnittstelle können beispielsweise Informationen über die aus dem Eingangssignal separierten Audiosignale einem Nutzer angezeigt werden. Der Nutzer kann dann eine Priorisierung einzelner der Audiosignale manuell vornehmen. Diese Nutzereingaben können zur Verarbeitung der Audiosignale herangezogen werden. Auch allgemeine Vorlieben des Nutzers, beispielsweise eine Unterdrückung von Umgebungsgeräuschen, kann mit Hilfe der Nutzerschnittstelle an die Signalverarbeitungseinrichtung übergeben werden und bei der Verarbeitung der Audiosignale berücksichtigt werden. Die Verarbeitungseinrichtung ist besonders gut individualisierbar.

[0031] Gemäß einem weiteren vorteilhaften Aspekt der Erfindung weist die Signalverarbeitungseinrichtung mindestens einen Datenspeicher zum Speichern bekannter Arten von Audiosignalen auf. Beispielsweise können Sprachmuster bestimmter Sprecher gespeichert werden. Der mindestens eine Datenspeicher erlaubt daher ein Archivieren von bereits bekannten Informationen über Audiosignale. Neu aus einem Eingangssignal separierte Audiosignale können mit diesem Datenspeicher abgeglichen werden. So können beispielsweise Audiosignale, die auf dem Nutzer bekannte Geräuschquellen, insbesondere dem Nutzer bekannte Sprecher, zurückgehen, identifiziert werden. Weiterhin kann die Relevanz der identifizierten Audiosignale für den Sprecher aus auf dem Datenspeicher hinterlegten Informationen ermittelt werden. Beispielsweise kann das Sprachmuster von Familienangehörigen eines Nutzers der Signalverarbeitungseinrichtung gespeichert werden, sodass auf diese zurückgehende Audiosignale für den Nutzer verstärkt werden.

[0032] Besonders bevorzugt kann die Signalverarbeitungseinrichtung die bekannten Arten von Audiosignalen auch extern, beispielsweise auf einem Cloud-Speicher, speichern. Dies hat den Vorteil, dass das Nutzerprofil nicht an eine bestimmte Signalverarbeitungseinrichtung gebunden ist. Der Nutzer kann beim Wechsel der Signalverarbeitungseinrichtung das auf ihn zugeschnittene Profil weiterverwenden.

[0033] Mit Hilfe der Nutzerschnittstelle kann der Nutzer insbesondere Einfluss auf die Anzahl und Art der separierten Audiosignale nehmen. Der Nutzer kann insbesondere eine automatische Anpassung des Systems überschreiben. Die Nutzereingaben können auch gespeichert werden und durch das System ausgewertet werden. Hierdurch kann das System automatisch anhand der früheren Nutzereingaben Vorlieben des Nutzers erkennen und sich hieran adaptiv anpassen.

[0034] Bevorzugt ist die Signalverarbeitungseinrichtung automatisch an nutzerspezifische Daten, Systemparameter, das Eingangssignal und/oder zuvor bereits separierte Audiosignale anpassbar. Insbesondere ist die Anzahl und Art der verwendeten zweiten neuronalen Netzwerke automatisch anpassbar. Hierdurch kann die variable Anzahl aus dem Eingangssignal separierter Audiosignale automatisch und adaptiv verändert werden. Das System ist lernfähig und passt sich an die Bedürfnisse des Nutzers an.

[0035] Als nutzerspezifische Daten stehen beispielsweise der Standort und/oder Bewegungsdaten des Nutzers zur Verfügung. So kann beispielsweise anhand des Standorts und des Bewegungsprofil des Nutzers ermittelt werden, dass dieser am Straßenverkehr teilnimmt. In diesem Fall kann ein zur Separierung von Verkehrslärm spezialisiertes zweites neuronales Netzwerk ausgewählt werden. Die für den Nutzer relevanten Audiosignale, beispielsweise ein sich näherndes Auto oder ein Hupen, werden so zuverlässig aus dem Eingangssignal separiert. Die nutzerspezifischen Daten können beispielsweise mit Hilfe entsprechender Sensoren ermittelt werden und an die Signalverarbeitungseinrichtung übergeben werden.

[0036] Besonders bevorzugt ist die Signalverarbeitungseinrichtung mit weiteren Sensoren verbunden und/oder weist weitere Sensoren auf, um nutzerspezifische Daten und/oder Systemparameter zu ermitteln.

[0037] Es ist eine weitere Aufgabe der Erfindung, ein System, insbesondere ein Hörgerätesystem, zur Verarbeitung von Audiosignalen zu verbessern. Diese Aufgabe ist gelöst durch ein System mit den in Anspruch 8 angegebenen Merkmalen.

[0038] Das System weist die erfindungsgemäße Signalverarbeitungseinrichtung auf. Zudem hat das System mindestens eine Aufzeichnungseinrichtung zum Aufzeichnen eines Eingangssignals und mindestens eine Wiedergabeeinrichtung zum Wiedergeben eines Ausgangssignals auf. Die mindestens ei-

ne Aufzeichnungseinrichtung ist über die Eingangsschnittstelle in datenübertragender Weise mit der Signalverarbeitungseinrichtung verbunden. Die mindestens eine Wiedergabeeinrichtung ist über die Ausgabeschnittstelle in datenübertragender Weise mit der Signalverarbeitungseinrichtung verbunden. Das System weist die oben diskutierten Vorteile der Signalverarbeitungseinrichtung auf. Das System ist besonders bevorzugt ein Hörgerätesystem. In diesem Fall können Hörgeräte vorgesehen sein, die die mindestens eine Aufzeichnungseinrichtung und die mindestens eine Wiedergabeeinrichtung umfassen.

[0039] Die Wiedergabeeinrichtung ist insbesondere ein Lautsprecher, bevorzugt ein Kopfhörer, besonders bevorzugt ein In-Ear-Kopfhörer, wie er in Hörgeräten verwendet wird.

[0040] Die mindestens eine Aufzeichnungseinrichtung ist insbesondere ein Mikrophon. Bevorzugt sind mehrere räumlich getrennte Mikrophone vorgesehen. Beispielsweise können zwei Hörgeräte mit je einem Mikrophon ausgestattet sein. Zusätzlich können weitere Mikrophone, beispielsweise Mikrophone an eine Mobilfunkgerät und/oder einer Armbanduhr, insbesondere einer Smartwatch, verwendet werden. Alternativ oder zusätzlich hierzu können auch weitere Mikrophone verwendet werden. Beispielsweise könne weitere Mikrophone mit der Signalverarbeitungseinrichtung, insbesondere mit einem die Signalverarbeitungseinrichtung umfassenden Mobilfunkgerät gekoppelt werden. Bevorzugt kann ein weiteres Mikrophon dazu ausgelegt sein Geräusche aus einem Umkreis von 360° aufzunehmen. Derartige zusätzliche Geräte können zudem auch für die Datenverbindung zwischen den Hörgräten und der Signalverarbeitungseinrichtung verwendet werden. Bereits vor Aufbereitung mit dem ersten neuronalen Netzwerk kann, eine räumliche Ortung der Audiosignale erfolgen. Hierdurch können bereits wesentliche Informationen über die Audiosignale erhalten werden. Dies verbessert die Genauigkeit des Systems.

[0041] Gemäß einem bevorzugten Aspekt des Systems ist die mindestens eine Signalverarbeitungseinrichtung als mobiles Gerät, insbesondere als Teil eines Mobilfunkgeräts ausgebildet. Dies gewährleistet eine hohe Flexibilität des Systems, insbesondere des Hörgerätesystems. Moderne Mobilfunkgeräte haben eine hohe Rechenleistung und Akkukapazität. Dies ermöglicht einen autarken Betrieb des Systems über längere Zeiträume. Zudem hat diese Ausführung den Vorteil, dass das System mit von einem Nutzer ohnehin mitgeführter Hardware realisiert werden kann. Zusätzliche Geräte sind nicht nötig.

[0042] Eine als Teil eines Mobilfunkgeräts ausgeführte Signalverarbeitungseinrichtung kann durch Komponenten des Mobilfunkgeräts realisiert sein. Besonders bevorzugt werden hierzu die normalen Hardwarekomponenten des Mobilfunkgeräts verwendet, indem eine Software, beispielsweise in Form einer App, auf dem Mobilfunkgerät ausgeführt wird. Beispielsweise können das mindestens eine erste neuronale Netzwerk und das mindestens eine zweite neuronale Netzwerk, insbesondere in Form eines gemeinsamen neuronalen Netzwerks, auf einem AI-Chip des Mobilfunkgeräts ausgeführt werden. In anderen Fällen, kann das Mobilfunkgerät speziell für die Signalverarbeitungseinrichtung ausgelegte Hardwarekomponenten umfassen.

[0043] Gemäß einem weiteren vorteilhaften Aspekt der Erfindung ist das System modular aufgebaut. Dies gewährleistet eine flexible Anpassung des Systems an die jeweiligen Nutzervorlieben. Einzelne Komponenten des Systems können, insbesondere bei Defekt, ausgetauscht werden. Beispielsweise können ein oder mehrere Hörgeräte mit einem beliebigen Mobilfunkgerät, auf dem die entsprechende Software installiert ist, kombiniert werden.

[0044] Es ist eine weitere Aufgabe der Erfindung, ein verbessertes Verfahren zur Verarbeitung von Audiosignalen bereitzustellen.

[0045] Diese Aufgabe wird gelöst durch ein Verfahren mit den in Anspruch 11 angegebenen Schritten. Zunächst wird die erfindungsgemäße Signalverarbeitungsrichtung bereitgestellt. Zudem wird ein Eingangssignal bereitgestellt. Dies kann beispielsweise über mindestens eine Aufzeichnungseinrichtung erfolgen. Das Eingangssignal wird zu der Signalverarbeitungseinrichtung über die Eingangsschnittstelle zugeführt. Hieraufhin wird das Eingangssignal mit Hilfe des mindestens einen ersten neuronalen Netzwerks aufbereitet. Mit Hilfe des mindestens einen neuronales Netzwerks, das sequentiell auf das mindestens eine erste neuronale Netzwerk folgt, werden einzelne oder mehrere Audiosignale aus dem aufbereiteten Eingangssignal separiert. Zu jedem der separierten Audiosignale wird ein Prioritätsparameter bestimmt. In Abhängigkeit von dem jeweiligen Prioritätsparameter wird jedes Audiosignal moduliert. Anschließend werden die modulierten Audiosignale zu einem Ausgangssignal kombiniert, das über die Ausgangsschnittstelle ausgegeben wird.

[0046] Durch die Separierung einzelner oder mehrerer der Audiosignale, können diese in dem Verfahren vorteilhafterweise getrennt moduliert werden. Dies ermöglicht eine unabhängige Anpassung der einzelnen Audiosignale, die an den jeweiligen Nutzer individuell angepasst werden kann. Der Prioritätsparameter ist bevorzugt kontinuierlich, sodass eine kontinuierliche Anpassung der Modulierung an die Relevanz der jeweiligen Audiosignale und/oder an die Vorlieben des Nutzers erfolgen kann. Beispielsweise kann der Prioritätsparameter zwischen 0 und 1 betragen. Die niedrigste Relevanz hätten dann Audiosignale mit

dem Prioritätsparameter 0, welche vollständig unterdrückt würden. Die höchste Priorität hätten Audiosignale mit dem Prioritätsparameter 1, was eine maximale Verstärkung des Audiosignals bewirken würde. Alternativ kann der Prioritätsparameter auch diskret sein, sodass die unterschiedlichen Audiosignale in verschiedene Klassen eingeteilt werden.

[0047] Gemäß einem vorteilhaften Aspekt des Verfahrens werden die separierten Audiosignale klassifiziert. Hierunter ist zu verstehen, dass die Audiosignale in verschiedene, der jeweiligen Art des Audiosignals entsprechende Gruppen eingeteilt werden. Bevorzugt ist das mindestens eine zweite neuronale Netzwerk an eine bestimmte Art von Audiosignalen angepasst, wie dies oben beschrieben ist. Ein derart angepasstes zweites neuronales Netzwerk separiert bevorzugt Audiosignale der jeweiligen Art aus dem Eingangssignal. Auf diese Weise wird durch das Separieren der Audiosignale mit Hilfe des mindestens einen neuronalen Netzwerks eine implizierte Klassifizierung des separierten Audiosignals vorgenommen. Die Klassifizierung kann jedoch auch nach der Separierung erfolgen, beispielsweise indem die Audiosignale asynchron analysiert und/oder mit weiteren nutzerspezifischen Daten kombiniert werden.

[0048] Besonders bevorzugt erfolgt nicht nur eine Klassifizierung sondern auch eine Identifizierung der Audiosignale. So kann beispielsweise nicht nur die Art des Audiosignals sondern auch eine bestimmte Quelle des Audiosignals erkannt werden. So kann zunächst mit Hilfe des mindestens einen zweiten neuronalen Netzwerks das Audiosignal implizit als gesprochene Sprache klassifiziert werden. Durch eine Analyse des separierten Audiosignals beispielsweise durch einen Abgleich mit auf einem Datenspeicher gespeicherten bekannten Audiosignalen, kann dann der jeweilige Sprecher identifiziert werden.

[0049] Gemäß einem weiteren vorteilhaften Aspekt des Verfahrens erfolgt die Auswahl des mindestens einen zweiten neuronalen Netzwerks aus einer zur Verfügung stehenden Menge von unterschiedlichen zweiten neuronalen Netzwerken anhand nutzerspezifischer Daten und/oder bereits separierter Audiosignale. Durch die Auswahl des mindestens einen zweiten neuronalen Netzwerks wird das Verfahren an das jeweilige Eingangssignal und darin enthaltenen Audiosignale noch besser angepasst. Als nutzerspezifische Daten stehen hierbei beispielsweise der Standort und/oder Bewegungsdaten des Nutzers zur Verfügung. So kann beispielsweise anhand des Standorts und des Bewegungsprofils des Nutzers ermittelt werden, dass dieser am Straßenverkehr teilnimmt. In diesem Fall kann ein zur Separierung von Verkehrslärm spezialisiertes zweites neuronales Netzwerk ausgewählt werden. Die für den Nutzer relevanten Audiosignale, beispielsweise ein sich näherndes Auto oder ein Hupen, werden so zuverlässig aus dem Eingangssignal separiert und können ihrer jeweiligen Relevanz entsprechend moduliert werden.

[0050] Die Auswahl des mindestens einen zweiten neuronalen Netzwerks kann zusätzlich oder alternativ anhand bereits separierter Audiosignale erfolgen. Beispielsweise kann ein separiertes Audiosignal als ein sich näherndes Kraftfahrzeug identifiziert werden. In diesem Fall kann ebenfalls das auf Verkehrslärm spezialisierte zweite neuronale Netzwerk ausgewählt werden, um zuverlässig Audiosignale, die auf verschiedene Kraftfahrzeuge zurückgehen, separieren können. Durch die Berücksichtigung bereits separierter Audiosignale bei der Auswahl des mindestens einen zweiten neuronalen Netzwerks ist das Verfahren selbstadaptiv.

[0051] Die Auswahl zweiter neuronaler Netzwerke kann zusätzlich oder alternativ anhand von Systemparametern erfolgen. Beispielhafte Systemparameter sind eine der Signalverarbeitungseinrichtung zur Verfügung stehende Rechenleistung und/oder der Akkuladestand, der der Signalverarbeitungseinrichtung noch zur Verfügung steht. Sinkt beispielsweise der verbleibende Akkuladestand unter einen vorbestimmten Grenzwert, kann die Anzahl der zweiten neuronalen Netzwerke verringert werden, um eine energiesparende Separation vorzunehmen. Alternativ können auch zweite neuronale Netzwerke mit weniger Outputs verwendet werden, um eine Separation mit geringerem Stromverbrauch zu ermöglichen. Die Anzahl der verwendeten zweiten neuronalen Netzwerke, insbesondere die Anzahl der aus dem Eingangssignal separierten Audiosignale, kann auch die jeweils der Signalverarbeitungseinrichtung zur Verfügung stehende Rechenleistung angepasst werden. Dies ist insbesondere von Vorteil, wenn die Signalverarbeitungseinrichtung Teil eines Mobilfunkgeräts ist. Beispielsweise kann ein Prozessor des Mobilfunkgeräts nicht nur zur Ausführung der zweiten neuronalen Netzwerke, sondern auch für anderweitige Rechenoptionen benutzt werden. Ist der Prozessor mit derartigen anderweitigen Rechenoperationen belegt, kann die Anzahl der zweiten neuronalen Netzwerke reduziert werden. Die Signalverarbeitungseinrichtung schränkt die sonstige Nutzung des Mobilfunkgeräts durch den Nutzer prinzipiell nicht ein.

[0052] Gemäß einem weiteren vorteilhaften Aspekt des Verfahrens erfolgt die Bestimmung der Prioritätsparameter asynchron zu weiteren Schritten des Verfahrens. Für die Bestimmung der Prioritätsparameter kann eine weitere Analyse der separierten Audiosignale nötig sein. Die asynchrone Bestimmung der Prioritätsparameter gewährleistet, dass die Bestimmung die Modulierung der Audiosignale und die Ausgabe des Ausgangssignals nicht verzögert. Die Modulierung der Audiosignale und die Ausgabe des Ausgangssignals können ohne Verzögerung erfol-

gen. Der Nutzer hört die modulierten Audiosignale in Echtzeit. Dies erhöht die Sicherheit und Genauigkeit bei der Durchführung des Verfahrens.

[0053] Aufgrund der asynchronen Bestimmung der Prioritätsparameter werden die Prioritätsparameter insbesondere schrittweise angepasst. Die Anpassung kann in festen Zeitabständen oder in dynamisch anpassbaren Zeitabständen erfolgen. Dies kann von der jeweiligen Nutzungssituation abhängen. Beispielsweise würde die Anpassung kurz getaktet erfolgen, wenn sich das Eingangssignal, insbesondere die hierin enthaltenen Audiosignale, und/oder die Prioritätsparameter oft und schnell ändern können, beispielsweise wenn der Nutzer am Straßenverkehr teilnimmt. Andererseits würde die Anpassung in längeren Taktungen erfolgen, wenn eine Änderung der Prioritätsparameter nicht zu erwarten ist, beispielsweise beim Fernsehen. Die Anpassung kann bis zu einmal alle 5 Millisekunden erfolgen. Die Anpassung kann auch nur einmal pro Sekunde erfolgen. Bevorzugt erfolgt die Anpassung nicht seltener als einmal alle zehn Minuten. Die Anpassungsrate kann zwischen einmal pro 5 Millisekunden und einmal pro 10 Minuten bevorzugt dynamisch variiert werden. Alternativ oder zusätzlich kann Anpassung auf das Erkennen bestimmter Signale erfolgen. Derartige Signale können ein Hupen oder ein Signalwort, wie beispielsweise „Hallo“, sein.

[0054] Besonders bevorzugt können auch weitere Schritte des Verfahrens anhand nutzerspezifischer Daten und/oder bereits separierter Audiosignale angepasst werden. Beispielsweise kann eine klassische Aufbereitung des Eingangssignals anhand der Anzahl der in dem Eingangssignal enthaltenen Audiosignale erfolgen.

[0055] Gemäß einem weiteren vorteilhaften Aspekt des Verfahrens erfolgt die Bestimmung der Prioritätsparameter abhängig von nutzerspezifischen Daten, Vorlieben des Nutzers und/oder einem Informationsgehalt der jeweiligen Audiosignale.

[0056] Über die nutzerspezifische Daten, beispielsweise einem Standort oder einem Bewegungsmuster des Nutzers, kann beispielsweise die Umgebung des Nutzers bestimmt werden. Je nach Umgebung werden verschiedene Prioritätsparameter bestimmt. Beispielsweise werden auf Kfz zurückgehende Audiosignale verstärkt, wenn der Nutzer am Straßenverkehr teilnimmt, wo die auditive Erfassung der Audiosignale von anderen Verkehrsteilnehmern sicherheitsrelevant ist. Wenn der Nutzer jedoch nicht am Straßenverkehr teilnimmt, beispielsweise in einem Straßencafe sitzt, werden diese Geräusche unterdrückt.

[0057] Durch Nutzerangaben können die Vorlieben des Nutzers berücksichtigt werden, beispielsweise werden bestimmte Personen besonders stark verstärkt, wohingegen andere, den Nutzer störende Geräusche, gezielt unterdrückt werden können. Besonders vorteilhaft ist die Bestimmung der Prioritätsparameter anhand des Informationsgehalts des jeweiligen Audiosignals. Beispielsweise kann ein Hupen oder ein Ausruf „Achtung“ verstärkt werden, um die Aufmerksamkeit des Nutzers, insbesondere in Gefahrensituationen, zu erregen. Um den Informationsgehalt des Audiosignals bestimmen zu können, kann das Audiosignal beispielsweise transkribiert und der transkribierte Inhalt ausgewertet werden.

[0058] Besonders bevorzugt werden die aus dem Eingangssignal separierten Audiosignale verbessert. Beispielsweise kann ein Rauschen, das auf ein schlechtes Mikrophon zurückzuführen ist, nicht mit den Audiosignalen aus dem Eingangssignal separiert werden. Die Audiosignale weisen somit eine hohe Qualität unabhängig von den verwendeten Mikrophonen auf. Zusätzlich oder alternativ hierzu können die Audiosignale nach der jeweiligen Separation auch aufbereitet werden. Hierzu können weitere neuronale Netzwerke und/oder Filter angewandt werden. Das aus den Audiosignalen zusammengesetzte Ausgangssignal hat eine hohe Qualität. Insbesondere bei Durchführung des Verfahrens in einem Hörgerätesystem kann der Nutzer die in dem Ausgangssignal enthaltenen Audiosignale auditiv einfach und zuverlässig erfassen. Audiosignale, die gesprochene Sprache enthalten, sind klar und deutlich verständlich.

[0059] Weitere Details, Merkmale und Vorteile der Erfindung ergeben sich aus der Beschreibung eines Ausführungsbeispiels anhand der Figuren. Es zeigen:

Fig. 1 eine schematische Darstellung eines Systems zur Verarbeitung von Audiosignalen,

Fig. 2 ein schematischer Verfahrensablauf beim Bearbeiten von Audiosignalen mit Hilfe des Systems gemäß **Fig. 1**,

Fig. 3 eine schematische Darstellung eines Vorbereitungsschritts des Verfahrens gemäß **Fig. 2**, und

Fig. 4 eine schematische Darstellung eines Separationsschritts des Verfahrens gemäß **Fig. 2**.

[0060] In **Fig. 1** ist schematisch ein System zur Verarbeitung von Audiosignalen in Form eines Hörgerätesystems **1** gezeigt. Das Hörgerätesystem **1** umfasst zwei Hörgeräte **2**, **3**, die am linken beziehungsweise rechten Ohr eines Nutzers getragen werden können. Zusätzlich weist das Hörgerätesystem **1** eine Signalverarbeitungseinrichtung **4** auf. Die Signalverarbeitungseinrichtung **4** ist Teil eines Mobilfunkgeräts **5**. Dies bedeutet, dass die Signalverarbeitungseinrichtung **4** durch Komponenten des Mobilfunkgeräts **5** realisiert ist. In dem dargestellten Ausführungs-

beispiel ist die Signalverarbeitungseinrichtung **4** realisiert, indem die Komponenten des Mobilfunkgeräts **5** eine entsprechende Software, die beispielsweise als App auf dem Mobilfunkgerät **5** installiert werden kann, ausführen. Die Signalverarbeitungseinrichtung **4** bedient sich also der Hardware des Mobilfunkgeräts **5**, wobei in **Fig. 1** die von der Signalverarbeitungseinrichtung verwendeten HardwareKomponenten durch eine gestrichelte Linie abgegrenzt dargestellt sind. Das Hörgerätesystem **1** ist modular ausgeführt. Unterschiedliche Mobilfunkgeräte können zur Realisierung der Signalverarbeitungsvorrichtung **4** verwendet werden. Auch kann nur eines der Hörgeräte **2**, **3** mit der Signalverarbeitungseinrichtung **4** gekoppelt werden.

[0061] In anderen, nicht dargestellten Ausführungsbeispielen können separate Hardwarekomponenten in einem Mobilfunkgerät zur Realisierung der Signalverarbeitungsvorrichtung **4** vorgesehen sein. In wiederum anderen, nicht dargestellten Ausführungsbeispielen wird die Signalverarbeitungseinrichtung **4** auf anderen mobilen Geräten, beispielsweise Smartwatches, realisiert. Auch ist möglich, dass die Signalverarbeitungseinrichtung **4** direkt in einem der Hörgeräte **2**, **3** integriert ist.

[0062] Die Hörgeräte **2**, **3** weisen jeweils ein Mikrofon **6** und einen Lautsprecher **7** auf. Die Hörgeräte **2**, **3** sind jeweils über eine drahtlose Datenverbindung **8** mit dem Mobilfunkgerät **5** verbunden. In dem dargestellten Ausführungsbeispiel ist die Datenverbindung **8** eine Bluetooth-Verbindung. In anderen Ausführungsbeispielen können auch andere Arten von Datenverbindungen verwendet werden. Die Datenverbindung kann insbesondere auch über zusätzliche Geräte hergestellt werden. Hierfür weisen das Mobilfunkgerät **5** sowie die Hörgeräte **2**, **3** jeweils eine Bluetooth-Antenne **9** auf. Die Signalverarbeitungseinrichtung **4** weist ein erstes neuronales Netzwerk **10** und eine Mehrzahl zweiter neuronaler Netzwerke **11** auf. In **Fig. 1** sind beispielhaft zwei zweite neuronale Netzwerke **11** dargestellt. Die Anzahl der zweiten neuronalen Netzwerke **11** kann jedoch variieren, wie dies im Folgenden noch beschrieben wird. Das erste neuronale Netzwerk **10** und die zweiten neuronalen Netzwerke **11** sind sequentiell angeordnet, d.h. dass ein Output des ersten neuronalen Netzwerks **10** als Input für die zweiten neuronalen Netzwerke **11** dient. Das erste neuronale Netzwerk **10** und die zweiten neuronalen Netzwerke **11** sind Teil eines gemeinsamen neuronalen Netzwerks, das mit Hilfe der Signalverarbeitungseinrichtung **4** ausgeführt wird. Wie oben bereits beschrieben, wird die Signalverarbeitungseinrichtung **4** durch Komponenten des Mobilfunkgeräts **5** realisiert. Die neuronalen Netzwerke **10**, **11** werden in dem dargestellten Ausführungsbeispiel daher auf einer Recheneinheit **12** des Mobilfunkgeräts **5** ausgeführt. Die Recheneinheit **12** des Mobilfunkgeräts **5** weist einen AI-Chip auf, wodurch die neuronalen Netzwerke **10**, **11** besonders effizient ausgeführt werden können. Der AI-Chip weist beispielsweise 2 Terraflops oder mehr auf.

[0063] Die Signalverarbeitungseinrichtung **4** weist zudem eine Eingangsschnittstelle **13** zum Empfangen eines Eingangssignals und eine Ausgangsschnittstelle **14** zur Ausgabestelle eines Ausgangssignals auf. Zudem ist ein Datenspeicher **15** vorgesehen, in welchem prozessrelevante Daten gespeichert werden können. Mit Hilfe einer weiteren Datenschnittstelle **16** können die im Datenspeicher **15** gespeicherten Daten auch auf einem externen Speicher **17** gespeichert werden. Als besonders geeignet hat sich für den externen Speicher **17** ein Cloud-Speicher erwiesen. Die Datenschnittstelle **16** kann insbesondere eine Mobilfunkdaten- oder W-LAN-Schnittstelle sein. Zudem weist die Signalverarbeitungseinrichtung **4** eine Nutzerschnittstelle **18** auf. Mit Hilfe der Nutzerschnittstelle **18** können Daten an einen Nutzer ausgegeben werden, in dem diese beispielsweise auf einem nicht dargestellten Display des Mobilfunkgeräts **5** angezeigt werden. Zudem können über die Nutzerschnittstelle **18** Nutzereingaben, beispielsweise über einen nicht dargestellten Touchscreen des Mobilfunkgeräts **5** an die Signalverarbeitungseinrichtung **4** übergeben werden.

[0064] Das Mobilfunkgerät **5** weist mindestens ein weiteres Mikrofon **19** auf, welches mit der Eingangsschnittstelle **13** verbunden ist. Zudem ist die Recheneinheit **12** mit weiteren Sensoren **20** des Mobilfunkgeräts **5** verbunden. So kann die Signalverarbeitungseinrichtung **4** beispielsweise auf mit Hilfe eines DPS-Sensors ermittelte Standortdaten und/oder mit Hilfe eines Bewegungssensors ermittelten Bewegungsdaten des Nutzers zugreifen.

[0065] Das erste neuronale Netzwerk **10** umfasst in dem Ausführungsbeispiel eine 1D Konvolutionsebene und zwei LSTM Ebenen mit 1024 bzw. 512 Einheiten. Das Eingangssignal kann so in eine neue, kompaktere Repräsentation gebracht werden. Skip Connections zwischen den Ebenen ermöglichen auch den Zugriff auf das originale Eingangssignal und auf alle Zwischenresultate. Die zweiten neuronalen Netzwerke **11** weisen zwei LSTM Ebenen mit 265 bzw. 128 Einheiten auf. An die LSTM Ebenen der zweiten neuronalen Netzwerke **11** schließen sich zwei vollständig verknüpften Ebenen mit 128 bzw. 64 Einheiten und eine 1D Konvolutionsebene an. In anderen Ausführungsbeispielen können die neuronalen Netzwerke **10**, **11** unterschiedliche Anzahlen von Ebenen und/oder Einheiten aufweisen oder gänzlich andere Strukturen aufweisen. Die verwendete Architektur der neuronalen Netzwerke ist für die Separation der Audiosignale aus dem Eingangssignal nicht wesentlich.

[0066] Die neuronalen Netzwerke **10**, **11** dienen zur Separierung einzelner Audiosignale aus einem Eingangssignal. Das erste neuronale Netzwerk **10** dient hierbei dazu, ein verschiedene Audiosignale umfassendes Eingangssignal derart aufzubereiten, dass die zweiten neuronalen Netzwerke eine effiziente Separierung von Audiosignalen aus dem Eingangssignal vornehmen können. Die Aufbereitung erfolgt unabhängig von der Form des jeweiligen Eingangssignals. Es wird daher unabhängig vom Eingangssignal immer das gleiche erste neuronale Netzwerk **10** verwendet. Dies ist besonders effizient. Das Eingangssignal umfasst die letzten Millisekunden von mithilfe der Mikrophone **6**, **19** kontinuierlich aufgezeichneten Audiodaten. Bei einer Rate von 16000 Samples pro Sekunde der Audiodaten umfasst das Eingangssignal etwa 128 Samples pro Kanal. Das Eingangssignal wird in Form eines 2-dimensionalen Tensors (Matrix) verarbeitet, wobei die Anzahl der Reihen die Anzahl Kanäle und die Anzahl der Zeilen die Anzahl an Samples repräsentiert. Das Signal wird mit einer Auflösung von 16 Bit verarbeitet, was die Effizienz erhöht, ohne die Sprachqualität signifikant zu beeinflussen. Das Eingangssignal wird in dem ersten neuronalen Netzwerk **11** durch den 1D convolutional Layer zunächst umgewandelt und mittels der LSTM Ebenen aufbereitetet

[0067] Die zweiten neuronalen Netzwerke **11** sind jeweils an die Erkennung und Separierung bestimmter Arten von Audiosignalen, beispielsweise gesprochene Sprache oder Verkehrslärm, angepasst. Die zweiten neuronalen Netzwerke **11** werden daher abhängig von den jeweiligen aus dem Eingangssignal zu separierenden Audiosignalen ausgewählt. Die Signalverarbeitungseinrichtung **4** weist hierfür eine Vielzahl an unterschiedliche Arten von Audiosignalen angepasste zweite neuronale Netzwerke auf. Die Anzahl und Zusammensetzung der zweiten neuronalen Netzwerke variiert daher mit dem jeweiligen Eingangssignal, wie dies später noch im Detail beschrieben wird.

[0068] Das erste neuronale Netzwerk **10** und die zweiten neuronalen Netzwerke **11** bilden zusammen ein gemeinsames Netzwerk. Das erste neuronale Netzwerk **10** bildet hierbei den Körper des gemeinsamen neuronalen Netzwerks, mit welchem immer wiederkehrende gleiche Aufgaben erledigt werden. Die zweiten neuronalen Netzwerke **11** bilden rotierende Köpfe des gemeinsamen neuronalen Netzwerks, die situationsabhängig ausgetauscht werden können. Somit ist eine besonders effiziente Kombination zwischen dem ersten neuronalen Netzwerk **10** und den zweiten neuronalen Netzwerken **11** geschaffen, ohne dass die Flexibilität der Separierung der Audiosignale eingeschränkt ist. Durch die Kombination der variablen zweiten neuronalen Netzwerke **11** mit dem ersten neuronalen Netzwerk **10** ist insbesondere eine Kontinuität bei der Separierung der Audiosignale gewährleistet. Ein Informationsverlust aufgrund des Wechsels einer oder mehrerer der zweiten neuronalen Netzwerke **11** ist vermieden, da Informationen im ersten neuronalen Netzwerk **10** gespeichert sind. Dies ist besonders vorteilhaft, da das Eingangssignal nur wenige Millisekunden umfasst. Mittels der LSTM Architektur können Informationen über einen längeren Zeitraum der aufgezeichneten Audiodaten in dem ersten neuronalen Netzwerk **11** gespeichert werden. Diese Informationen können dann auch nach einem Austausch von zweiten neuronalen Netzwerken an die neuen zweiten neuronalen Netzwerke übergeben werden. Die neuen zweiten neuronalen Netzwerke können anhand der gespeicherten Informationen optimal initiiert werden.

[0069] Mit Bezug auf die **Fig. 2** bis **Fig. 4** wird anhand eines konkreten Beispiels die Separation einzelner Audiosignale im Detail beschrieben. Hierzu sind die einzelnen hierfür nötigen Schritte unabhängig von den hierfür verwendbaren Hardwarekomponenten in Funktionsschritte unterteilt, wie sie in **Fig. 2** dargestellt sind.

[0070] In der in **Fig. 2** dargestellten Situation ist der Nutzer des Hörgerätesystems **1** mit verschiedenen Geräuschquellen konfrontiert. Beispielhaft dargestellt ist ein Sprecher **A**, der sich mit dem Nutzer des Hörgerätesystems **1** unterhält. Weiterhin sind zwei sich unterhaltende Passanten **B1** und **B2** in Hörweite. Zudem sind ein Auto **C** und ein Helikopter **D** zu hören.

[0071] Die von den Geräuschquellen emittierten Geräusche **G** werden in einem Aufzeichnungsschritt **21** mit Hilfe der Mikrophone **6** der Hörgeräte **3** und des Mikrophons **19** des Mobilfunkgeräts **5** aufgezeichnet und digitalisiert. Mit Hilfe der Datenverbindung **8** werden die mit den Mikrophonen **6** aufgezeichneten und digitalisierten Geräusche an das Mobilfunkgerät **5** übermittelt. Die mit Hilfe der Mikrophone **6** und des Mikrophons **19** ermittelten Geräusche werden zu einem Kanal **E1**, **E2**, **E3** je Mikrophon **6**, **19** enthaltenden Eingangssignal E kombiniert und an die Eingangsschnittstelle **13** der Signalverarbeitungseinrichtung **4** übermittelt. Die Signalverarbeitungseinrichtung **4** bedient sich in dem dargestellten Ausführungsbeispiel einiger der Komponenten des Mobilfunkgeräts **5**, wobei die von der Signalverarbeitungseinrichtung verwendeten Komponenten durch eine gestrichelte Linie abgegrenzt sind. Das Eingangssignal E wird in einem Vorbereitungsschritt **22** aufbereitet. Der Vorbereitungsschritt **22** ist in **Fig. 3** im Detail gezeigt. Das Eingangssignal E enthält die den unterschiedlichen Mikrophonen **6**, **19** entsprechenden Kanäle **E1**, **E2**, **E3**. Zunächst wird das Eingangssignal E während einer klassischen Vorbereitung **23** aufbereitet. Beispielsweise kann aufgrund der unterschiedlichen Kanäle **E1**, **E2**, **E3** des Eingangssignals E eine Vorklassifizierung der unterschiedlichen Audiosignale erfolgen, beispielsweise indem die relative Positi-

on anhand der unterschiedlichen, bekannten Positionen der Mikrofone ermittelt wird. In der klassischen Aufbereitung **23** werden die einzelnen Kanäle **E1**, **E2**, **E3** des Eingangssignals E zudem normalisiert und zu einem einheitlichen, alle Kanäle **E1**, **E2**, **E3** umfassenden Eingangssignal E' zusammengefasst. Das gemeinsame Eingangssignal E' stellt eine vereinheitlichte Repräsentation aller aufgezeichneten Geräusche dar. Das gemeinsame Eingangssignal E' dient als Eingangssignal für das erste neuronale Netzwerk **10**. Das erste neuronale Netzwerk **10** ist in **Fig. 3** rein schematisch dargestellt. Das erste neuronale Netzwerk **10** bereitet das gemeinsame Eingangssignal E' für die weitere Separierung der einzelnen Audiosignale in E' auf. Das aufbereitete Eingangssignal wird in Form eines Tensors **T** von dem ersten neuronalen Netzwerk **10** ausgegeben. Mit der Ausgabe des Tensors **T** endet der Vorbereitungsschritt **22**. Die einzelnen Kanäle **E1**, **E2**, **E3** enthalten verschiedene Mischungen der Audiosignale der jeweiligen Geräuschquellen.

[0072] An den Vorbereitungsschritt **22** schließt sich ein Separationsschritt **24** an. Der Separationsschritt **24** ist im Detail in **Fig. 4** gezeigt. In Separationsschritt **24** erfolgt eine Separierung einzelner Audiosignale mit Hilfe der zweiten neuronalen Netzwerke **11**. Die zweiten neuronalen Netzwerke **11** sind in **Fig. 4** rein schematisch dargestellt. In dem gezeigten Ausführungsbeispiel werden hierzu zwei unterschiedliche zweite neuronale Netzwerke **11** verwendet. Hierzu wird der im Vorbereitungsschritt **22** mit Hilfe des ersten neuronalen Netzwerks **10** ermittelte Tensor **T** zunächst der Anzahl der zweiten neuronalen Netzwerke **11** entsprechend vervielfältigt in einem Vervielfältigungsschritt **25**. Hierdurch ist sichergestellt, dass alle der im Separationsschritt **24** verwendeten zweiten neuronalen Netzwerke **11** den gleichen Input, nämlich den Tensor **T** erhalten. Mit anderen Worten, der durch das erste neuronale Netzwerk **10** ermittelte Tensor **T** wird an alle zweiten neuronalen Netzwerke **11** übergeben. Die beiden im Separationsschritt **24** verwendeten zweiten neuronalen Netzwerke **11** sind an unterschiedliche Arten von Audiosignalen angepasst. Jedes der zweiten neuronalen Netzwerke gibt eine bestimmte Anzahl von Output aus. Die Anzahl von Output ist für jedes der zweiten neuronalen Netzwerke **11** konstant, kann jedoch für verschiedene der zweiten neuronalen Netzwerke **11** unterschiedlich sein. Im dargestellten Ausführungsbeispiel geben die beiden verwendeten zweiten neuronalen Netzwerke **11** jeweils drei Outputs aus.

[0073] Das in **Fig. 4** oberhalb dargestellte zweite neuronale Netzwerk **11** ist auf die Erkennung und Separierung von gesprochener Sprache spezialisiert. Dieses Netzwerk wird die in dem Tensor **T** enthaltenen Audiosignale a, b1, b2 des Gesprächspartners **A** beziehungsweise der weiteren Passanten **B1** und **B2** erkennen und aus dem Tensor **T** separieren. Die Outputs des oberhalb dargestellten zweiten neuronalen Netzwerks **11** entsprechen daher den Audiosignalen des Gesprächspartners **A** sowie der weiteren Passanten **B1** und **B2**.

[0074] Das in **Fig. 4** unterhalb dargestellte zweite neuronale Netzwerk **11** ist auf die Erkennung und Separierung von Verkehrslärm spezialisiert. Dieses wird die in dem Tensor **T** enthaltenen Audiosignale c und d des Autos **C** sowie des Hubschraubers **D** erkennen und als Audiosignale ausgeben. Aufgrund der fixen Anzahl von Outputs je zweiten neuronalen Netzwerk **11** wird das unterhalb dargestellte zweite neuronale Netzwerk **11** auch ein weiteres in dem Tensor **T** enthaltenes Audiosignal separieren und ausgeben. In dem dargestellten Beispiel ist dies die durch den Passanten B 1 erzeugte gesprochene Sprache.

[0075] Da die unterschiedlichen zweiten neuronalen Netzwerke **11** an unterschiedliche Arten von Audiosignalen angepasst sind, separieren diese bevorzugt die jeweiligen Arten von Audiosignalen, beispielsweise Audiosignale bestimmter Arten von Geräuschquellen, wie beispielsweise Autos oder Sprecher. Durch die Separation mit Hilfe der zweiten neuronalen Netzwerke **11** werden die Audiosignale daher gemäß ihrer jeweiligen Art, insbesondere ihres jeweiligen Ursprungs klassifiziert. Die Separierung der Audiosignale mit Hilfe der zweiten neuronalen Netzwerke **11** lässt daher schon Rückschlüsse auf die Art der jeweiligen Audiosignale zu.

[0076] Mithilfe der zweiten neuronalen Netzwerke **11** werden die Audiosignale nicht nur separiert, sondern auch verbessert. Ein Rauschen, das beispielsweise auf ein schlechtes Mikrophon **6**, **19** zurückgeht, wird nicht zusammen mit den Audiosignalen aus dem Eingangssignal separiert. Die Signalverarbeitungseinrichtung ermöglicht eine hohe Qualität der Audiosignale unabhängig von den verwendeten Mikrophonen **6**, **19**.
Mit der Ausgabe der einzelnen Audiosignale endet der Separationsschritt **24**.

[0077] Vor der Weiterverarbeitung der separierten Audiosignale werden diese in einem Zusammenführschritt **26** auf Duplikate hin überprüft. Sollten einzelne der Outputs der zweiten neuronalen Netzwerke **11** dasselbe Audiosignal enthalten, werden diese Outputs zusammengeführt. Im dargestellten Ausführungsbeispiel betrifft dies die Sprache des Passanten **B1**, die in zwei Outputs der zweiten neuronalen Netzwerke **11** enthalten ist. Nach dem Zusammenführschritt **26** ist jedes der Audiosignale einmalig.

[0078] An den Zusammenführschritt **26** schließt sich ein Modulationsschritt **27** an. In dem Modulationsschritt **27** werden die Audiosignale moduliert, d.h. die einzelnen Audiosignale werden verstärkt oder unterdrückt. Die Entscheidung, welches der Audiosigna-

le verstärkt oder unterdrückt wird erfolgt mit Hilfe eines Prioritätsparameters, der jedem der Audiosignale zugeordnet wird. Der Prioritätsparameter kann einen Wert zwischen 0, was einer maximalen Unterdrückung des jeweiligen Audiosignals entspricht, und 1, was einer maximalen Verstärkung des jeweiligen Audiosignals entspricht, betragen.

[0079] Die Zuordnung des Prioritätsparameters erfolgt asynchron zu weiteren Schritten des Verfahrens, in einem asynchronen Klassifizierungsschritt **28**, wie dies nachfolgend noch beschrieben wird. Die asynchrone Bestimmung des Prioritätsparameters zu jedem der separierten Audiosignale hat den Vorteil, dass die Modulierung im Modulierungsschritt **27** ohne Verzögerung erfolgt. Die im Separationsschritt **24** separierten Audiosignale könne somit ohne Verzögerung anhand des jeweiligen Prioritätsparameters moduliert werden. Die modulierten Audiosignale werden in einem Ausgabeschritt **29** zu einem Ausgabesignal O zusammengefasst und mit Hilfe der Ausgabeschnittstelle **14** der Signalverarbeitungseinheit **4** ausgegeben. Im dargestellten Ausführungsbeispiel bedeutet dies, dass das Ausgabesignal O mit Hilfe der Ausgabeschnittstelle **14** an die Bluetooth-Antenne **9** des Mobilfunkgeräts **5** übergeben wird und von dort an die Hörgeräte **2**, **3** übertragen wird. Die Hörgeräte **2**, **3** geben das Ausgabesignal O mit Hilfe der Lautsprecher **7** wieder. Um ein Stereosignal zu erzeugen, enthält das Ausgabesignal O zwei Kanäle, die ein anhand der in der klassischen Aufbereitung **23** bestimmten Richtungen der Geräuschquellen entsprechendes Stereosignal bilden. In dem Wiedergabeschritt **30** werden die im Ausgabesignal O enthaltenen Kanäle mit Hilfe der entsprechenden Lautsprecher **7** wiedergegeben und sind für den Nutzer hörbar.

[0080] In anderen Ausführungsbeispielen wird das Ausgabesignal O als Monosignal mit nur einem Kanal ausgegeben. Dieses Ausgabesignal ist besonders effizient und praktikabel.

[0081] Im Folgenden wird beispielhaft die Zuordnung des Prioritätsparameters beschrieben. Die Zuordnung des Prioritätsparameters erfolgt in dem asynchronen Klassifizierungsschritt **28**. Der Prioritätsparameter wird anhand nutzerspezifischer Vorgaben, weiterer nutzerspezifischer Daten und/oder einem Abgleich mit bereits bekannten Audiosignalen ermittelt. Hierzu können beispielsweise in einem Sensorausleseschritt **31** Sensordaten der Sensoren **20** des Mobilfunkgeräts **5** ermittelt werden. Zudem können in einem Nutzereingabenausleseschritt **32** Nutzereingaben via der Nutzerschnittstelle **18** ausgelesen werden. In einem Datenabgleichschritt **33** können über die Audiosignale ermittelte Daten mit bereits auf dem internen Datenspeicher **15** und/oder dem externen Speicher **17** gespeicherten Informationen über bekannte Audiosignale abgeglichen werden.

[0082] In der in **Fig. 2** dargestellten Situation würde sich die Bestimmung der den einzelnen Audiosignalen zugeordneten Prioritätsparameter beispielsweise wie folgt gestalten:

> Der Nutzer des Hörgerätesystems **1** befindet sich beispielsweise schon in einem aktiven Gespräch mit dem Gesprächspartner **A**. In dem asynchronen Klassifizierungsschritt **28** wird das zugehörige Audiosignal a als gesprochene Sprache erkannt und kann mit einem für den Gesprächspartner **A** typischen, bereits bekannten und in dem Datenspeicher **15** hinterlegten Sprachmuster abgeglichen werden. Das Audiosignal a wird als dem Gesprächspartner zugehörig identifiziert und aufgrund dessen Relevanz für den Nutzer des Hörgerätesystems **1** als wichtig eingestuft. Dem Audiosignal a wird daher ein hoher Prioritätsparameter zugeordnet. Auch die den beiden Passanten B 1, B2 zugehörigen Audiosignale b1, b2 werden während des asynchronen Klassifizierschritts **28** als gesprochene Sprache erkannt. Jedoch sind die Passanten **B1**, **B2** dem Nutzer des Hörgerätesystems **1** nicht bekannt. Ein Abgleich mit bekannten, in dem Datenspeicher **15** gespeicherten Sprachmustern schlägt fehl. In der Folge wird den Audiosignalen b1, b2 ein geringer Prioritätsparameter zugeordnet, sodass eine Unterdrückung dieser Audiosignale erfolgt. Schaltet sich jedoch einer der beiden Passanten in das Gespräch mit dem Nutzer des Hörgerätesystems **1** ein, kann eine Reevaluierung dessen Audiosignals erfolgen. Dies kann beispielsweise automatisch geschehen, indem die Gesprächsteilnahme erkannt wird. Hierzu kann die Signalverarbeitungseinrichtung **4** Signalworte, wie beispielsweise „Hallo“ oder „Entschuldigung“ und/oder Sprechpausen ausgewertet werden. Zudem kann ein Transkript der erkannten Sprachsignale erstellt und inhaltlich ausgewertet werden. Die Signalverarbeitungseinrichtung **4** ist lernfähig und passt sich den Bedürfnissen des Nutzers automatisch an. Zusätzlich kann auch der Nutzer des Hörgerätesystems über eine Eingabe am Mobilfunkgerät **5**, die im Nutzereingabeausleseschritt **32** ausgelesen wird, dem jeweiligen Passanten einen höheren Prioritätsparameter zuordnen. Dies kann beispielsweise dadurch geschehen, dass die einzelnen separierten Audiosignale auf einem Display des Mobilfunkgeräts dem Nutzer angezeigt werden. Der Nutzer kann dann die jeweiligen bevorzugt zu behandelnden Audiosignale durch Touch-Eingabe auswählen. Die Nutzereingabe kann die automatische Anpassung des Systems überschreiben. Das Sprachmuster des entsprechenden Passanten kann dann als bekannte Audioquelle in dem Datenabgleichschritt **33** auf dem Datenspeicher **15** hinterlegt werden.

[0083] Das Audiosignal c des Autos **C** wird als sich in der Nähe des Nutzers des Hörgerätesystems **1** bewegendes Kraftfahrzeug erkannt. Je nachdem, welche weiteren Daten über den Standort und/oder die Bewegung des Nutzers mit Hilfe des Sensorauslese-schritts **31** ermittelt werden, kann der dem Audiosignal c zugeordnete Prioritätsparameter variieren. Ergibt die Standortabfrage und das Bewegungsmuster beispielsweise, dass der Nutzer in einem Straßencafe sitzt, hat das Audiosignal c des Autos **C** für den Nutzer in der Regel keinerlei Bedeutung. Diesem wird daher ein niedriger Prioritätsparameter zugeordnet. Bewegt sich der Nutzer jedoch im Straßenverkehr, ist das auditive Erfassen des sich bewegenden Fahrzeugs relevant für die sichere Beteiligung am Straßenverkehr. In diesem Fall wird dem Audiosignal c ein höherer Prioritätsparameter zugeordnet, sodass der Nutzer das sich nähernde Kraftfahrzeug erfassen kann.

[0084] Anders verhält sich dies für den Hubschrauber **D**. Dessen Audiosignal d ist in der Regel für die Sicherheit der Teilnahme am Straßenverkehr irrelevant. Dem Audiosignal d wird daher in dem asynchronen Klassifizierungsschritt **28** ein niedriger Prioritätsparameter zugeordnet. Jedoch kann auch hier der Nutzer durch entsprechende Nutzereingaben eine Anpassung des Prioritätsparameters bewirken.

[0085] Die Identifizierung der Audiosignale im asynchronen Klassifizierungsschritt **28** wird nicht nur zur Festlegung des Prioritätsparameters der einzelnen Audiosignale verwendet. Die im asynchronen Klassifizierungsschritt **28** gewonnenen Informationen über die Audiosignale werden auch zur Verbesserung deren Aufbereitung im Vorbereitungsschritt **22** und deren Separierung im Separationsschritt **24** herangezogen. Hierzu ist der asynchrone Klassifizierungsschritt **28** über eine Aufbereitungs-Feedbackschleife **34** mit dem Vorbereitungsschritt **22** gekoppelt. Über die Aufbereitungs-Feedbackschleife **34** werden in dem asynchronen Klassifizierungsschritt **28** gewonnene Informationen an den Vorbereitungsschritt **22** nachfolgend detektierter Eingangssignale übergeben. Diese Informationen betreffen die Umgebung des Nutzers des Hörgerätesystems **1** sowie die Anzahl und Qualität der zuvor separierten Audiosignale. Anhand dieser Informationen kann die klassische Aufbereitung **23** angepasst werden, beispielsweise indem die Normierung des Eingangssignals an die Anzahl der Audiosignale angepasst wird.

[0086] Mit Hilfe einer Separations-Feedbackschleife **35** werden die in dem asynchronen Klassifizierungsschritt **28** über die Audiosignale ermittelten Informationen an den Separationsschritt **24** nachfolgend aufgezeichneter Eingangssignale E übermittelt. Wie oben bereits erwähnt, sind die für den Separationsschritt **24** verwendeten zweiten neuronalen Netzwerke **11** austauschbar. Dies bedeutet, dass eine Vielzahl unterschiedlich konfigurierter beziehungsweise unterschiedlich spezialisierter zweiter neuronaler Netzwerke **11** für den Separationsschritt **24** verwendet werden können. Jedes der unterschiedlichen zur Verfügung stehenden zweiten neuronalen Netzwerke **11** ist an andere Arten von Audiosignalen angepasst. Mit Hilfe der über die Separations-Feedbackschleife **35** übermittelten Information kann in einem Netzwerkauswahlschritt **36** eine Auswahl der für den Separationsschritt **24** zur verwendeten zweiten neuronalen Netzwerke **11** erfolgen. Mit dem Netzwerkauswahlschritt **36** können alle oder einzelne der zweiten neuronalen Netzwerke, die für den Separationsschritt **24** verwendet werden, ausgetauscht werden. Zudem kann die Anzahl verwendeter zweiter neuronaler Netzwerke **11** variiert werden. Mit Hilfe der Separations-Feedbackschleife **35** kann beispielsweise die Anzahl der nach dem Zusammenführschritt **26** verbleibenden Audiosignale an den Vorbereitungsschritt **24** übermittelt werden. Da jedes der unterschiedlichen zweiten neuronalen Netzwerke **11** eine feste Anzahl von Outputs, d.h. eine feste Anzahl von einzelnen Audiosignalen, ausgibt, kann mit Hilfe der Information über die Anzahl der Netzwerke die Anzahl der für den Separationsschritt **24** verwendeten zweiten neuronalen Netzwerke **11** angepasst werden. Beispielsweise ist es möglich, dass weitere Geräuschquellen, beispielsweise Straßenbahnen oder weitere Passanten zu dem Eingangssignal beitragen, was eine Erhöhung der Anzahl der für den Separationsschritt **24** verwendeten zweiten neuronalen Netzwerke **11** bedingen kann. Zudem kann im Netzwerkauswahlschritt **36** die Anzahl der zur Separation der Audiosignale verwendeten zweiten neuronalen Netzwerke **11** an Parameter des Mobilfunkgeräts **5** angepasst werden. Sinkt beispielsweise dessen Akkuladestand unter einen vorbestimmten Grenzwert, kann die Anzahl der zweiten neuronalen Netzwerke **11** verringert werden, um eine energiesparende Separation vorzunehmen. Befindet sich der Nutzer allerdings in einer Situation mit vielen verschiedenen Audiosignalen und möchte er eine möglichst genaue Separierung haben, kann er die vorgenommene Reduktion der Anzahl der zweiten neuronalen Netzwerke **11** durch eine entsprechende Eingabe am Mobilfunkgerät **5** rückgängig machen. Im Netzwerkauswahlschritt **36** kann die Anzahl der verwendeten zweiten neuronalen Netzwerke **11** auch an die jeweils zur Verfügung stehende Rechenleistung angepasst werden. Beispielsweise kann die Recheneinheit **12** des Mobilfunkgeräts **5** mit anderweitigen Rechenoperationen belegt sein. Sodass die Anzahl der zweiten neuronalen Netzwerke **11** reduziert wird. Hierdurch ist gewährleistet, dass die Signalverarbeitungseinrichtung die sonstige Nutzung des Mobilfunkgeräts **5** durch den Nutzer nicht einschränkt.

[0087] Zudem kann im asynchronen Klassifizierungsschritt **28** auch die Qualität der Separation überprüft werden und die Auswahl der zweiten neurona-

len Netzwerke **11** mittels der Separations-Feedbackschleife **35** an die ermittelte Qualität angepasst werden. Zur Ermittlung der Qualität kann über einen längeren Zeitraum die Lautstärke einzelner der separierten Audiosignale gemessen werden. Dies kann mithilfe des quadratischen Mittelwerts (auch RMS vom englischen „root mean square“ genannt) und/oder über andere Charakteristika, wie zum Beispiel die maximale Lautstärke des Audiosignals erfolgen.

[0088] Des Weiteren kann eine Auswahl der zweiten neuronalen Netzwerke **11** anhand der Klassifizierung der Audiosignale im asynchronen Klassifizierungsschritt **28** erfolgen. Somit ist sichergestellt, dass für die Separation von Audiosignalen nachfolgender Eingangssignale jeweils das optimal an die jeweiligen Audiosignale angepasste zweite neuronale Netzwerk **11** verwendet wird. Indem in **Fig. 2** dargestellten Ausführungsbeispiel ist beispielsweise möglich, dass der Nutzer des Hörgerätesystems **1** einen Bahnhof betritt. Wird dies erkannt, beispielsweise anhand von der Auswertung von GPS-Daten, kann das auf Straßenverkehr spezialisierte zweite neuronale Netzwerk **11** gegen ein auf Bahnhofsgeräusche spezialisiertes zweites neuronales Netzwerk **11** ausgetauscht werden. Mit dem auf Bahnhofsgeräusche spezialisiertem zweiten neuronalen Netzwerk **11** können beispielsweise die Audiosignale einfahrender Züge unterdrückt werden, während Bahnhofsdurchsagen, beispielsweise die Verspätung eines Zuges betreffend, verstärkt werden.

[0089] Die Aufbereitungs-Feedbackschleife **34** und die Separations-Feedbackschleife **35** stellen sicher, dass sich die Signalverarbeitungseinrichtung an die jeweilige Geräuschkulisse und Umgebung des Nutzers des Hörgerätesystems **1** anpasst. Die Signalverarbeitungsanlage ist adaptiv.

[0090] Die Klassifizierung der Audiosignale im asynchronen Klassifizierungsschritt **28**, insbesondere deren Abgleich mit weiteren Sensordaten und/oder mit auf dem Datenspeicher **15** gespeicherten Informationen erfolgt asynchron zu den weiteren Schritten des Verfahrens. Hierdurch ist gewährleistet, dass die Separierung der Audiosignale im Separationsschritt **24** und die Modulierung der Audiosignale im Modulationsschritt in Echtzeit erfolgt, während die Klassifizierung im asynchronen Klassifizierungsschritt **28** abhängig von der Komplexität der Audiosignale und der weiteren Daten über einen gewissen Zeitraum erfolgt. Beispielsweise muss zum Abgleich eines Sprachmusters eines Sprechers mit gespeichertem Sprachmustern zunächst eine gewisse Sequenz des Sprachsignals aufgezeichnet und analysiert werden. Die Anpassung der Prioritätsparameter sowie des Vorbereitungsschritts **22** und des Separationsschritts **24** erfolgt dann schrittweise. Die Häufigkeit der Anpassung kann von den Hardwarekomponenten des Mobilfunkgeräts **5** und/oder von den die Umgebung betreffenden Umständen abhängen. So ist beispielsweise eine Anpassung der Prioritätsparameter im Straßenverkehr, in welchem sich die Geräuschkulisse oft ändern kann, mit einer wesentlich höheren Rate notwendig, als beispielsweise beim Fernsehen. Die Anpassung kann bis zu einmal alle 5 Millisekunden erfolgen. Die Anpassung erfolgt nicht seltener als einmal alle 10 Minuten. Zwischen diesen Eckparametern, kann die Anpassungsrate dynamisch variiert werden.

[0091] In dem beschriebenen Ausführungsbeispiel werden die Prioritätsparameter zu jedem Audiosignal stufenlos ermittelt. Somit ist eine stufenlose Abschätzung der Relevanz der einzelnen Audiosignale abhängig von den jeweiligen Umständen möglich. In anderen Ausführungsbeispielen kann der Prioritätsparameter auch eine Einteilung der einzelnen Audiosignale in unterschiedliche diskrete Klassen ermöglichen.

[0092] In dem Ausführungsbeispiel haben die jeweiligen zweiten neuronalen Netzwerke **11** jeweils eine bestimmte Anzahl an Outputs. Je Output wird ein aus dem Eingangssignal E separiertes Audiosignal ausgegeben. So gibt beispielsweise das in **Fig. 4** unterhalb dargestellte zweite neuronale Netzwerk **11**, das auf die Separierung spezialisiert ist, auch das Sprachsignal b1 des Passanten **B1** aus. In anderen Ausführungsbeispielen sind die zweiten neuronalen Netzwerke derart trainiert, dass diese jeweils nur Audiosignale der Art ausgeben, auf welche sie spezialisiert sind. In der in **Fig. 2** beispielhaft dargestellten Situation würde ein auf Verkehrslärm spezialisiertes zweites neuronale Netzwerk nicht das Sprachsignal b1 sondern nur die Audiosignale d, c des Hubschraubers **D** und des Autos **C** ausgeben. Der nicht belegte, überschüssige dritte Output würde dann ein leeres Signal ausgeben, was einem Audiosignal entspricht, das kein Geräusch bzw. Stille enthält. Die Anzahl leerer Signale, die von den zur Separierung verwendeten zweiten neuronalen Netzwerken erzeugt wird, dient der Signalverarbeitungseinrichtung als Maß für die Anzahl der im Eingangssignal enthaltenen Audiosignale. Erkennt die Signalverarbeitungseinrichtung, dass viele Outputs der zur Separierung verwendeten zweiten neuronalen Netzwerke leere Signale enthalten, kann die Anzahl der zur Separierung verwendeten zweiten neuronalen Netzwerke reduziert werden. So ist beispielsweise möglich, dass der Nutzer des Hörgerätesystems von der Straße in ein Haus tritt. Das zur Erkennung von Verkehrslärm verwendete zweite neuronale Netzwerk würde dann nur leere Signale ausgeben und könnte deaktiviert werden. Dies ermöglicht eine effiziente und stromsparende Separierung der Audiosignale.

[0093] In wiederum anderen Ausführungsbeispielen enthält ein Output jedes zweiten neuronalen Netzwerks ein Restsignal, das dem Eingangssignal ab-

züglich der mit Hilfe des jeweiligen zweiten neuronalen Netzwerks separierten Audiosignale enthält. Das Restsignal entspricht also jeweils der Summe aller nicht mit Hilfe des jeweiligen zweiten neuronalen Netzwerks aus dem Eingangssignal separierter Audiosignale. In der in **Fig. 2** beispielhaft dargestellten Situation würde ein auf Verkehrslärm spezialisiertes zweites neuronales Netzwerk daher die Audiosignale c, d des Autos **C** und des Helikopters **D** ausgeben. Das Restsignal würde dann die Sprachsignale a, b1, b2 des Gesprächspartner **A** sowie der Passanten **B1**, **B2** umfassen. Die jeweiligen Restsignale dienen der Signalverarbeitungseinrichtung als Maß für nicht separierte Audiosignale. Umfassen die Restsignale noch eine große Anzahl von Informationen, erhöht die Signalverarbeitungseinrichtung die Anzahl der zweiten neuronalen Netzwerke, wodurch eine größere Anzahl von Audiosignalen aus dem Eingangssignal separiert wird. Die Erhöhung der Anzahl der verwendeten zweiten neuronalen Netzwerke kann auch in diesem Fall durch die Erkennung weitere Systemparameter, wie beispielsweise einen niedrigen Akkustand, verhindert werden.

Patentansprüche

1. Signalverarbeitungseinrichtung zur Verarbeitung von Audiosignalen, mit
1.1. einer Eingangsschnittstelle (13) zum Empfangen eines Eingangssignals (E),
1.2. mindestens einem ersten neuronalen Netzwerk (10) zum Aufbereiten des Eingangssignals (E),
1.3. mindestens einem zweiten neuronalen Netzwerk (11) zum Separieren eines oder mehrerer Audiosignale (a, b1, b2, c, d) aus dem Eingangssignal (E), und
1.4. einer Ausgangsschnittstelle (14) zum Ausgeben eines Ausgangssignals (O),
1.5. wobei das mindestens eine erste neuronale Netzwerk (10) und das mindestens eine zweite neuronale Netzwerk (11) sequentiell angeordnet sind.

2. Signalverarbeitungseinrichtung nach Anspruch 1, **gekennzeichnet durch** eine Mehrzahl zweiter neuronaler Netzwerke (11), wobei jedes der zweiten neuronalen Netzwerke (1) an eine bestimmte Art von Audiosignalen (a, b1, b2, c, d) angepasst ist.

3. Signalverarbeitungseinrichtung nach einem der Ansprüche 1 oder 2, **dadurch gekennzeichnet**, dass mindestens zwei zweite neuronale Netzwerke (11) parallel zur Separierung der Audiosignale (a, b1, b2, c, d) aus dem Eingangssignal (E) verwendet werden.

4. Signalverarbeitungseinrichtung nach einem der vorgenannten Ansprüche, **dadurch gekennzeichnet**, dass das mindestens eine zweite neuronale Netzwerk (11) austauschbar ist.

5. Signalverarbeitungseinrichtung nach einem der vorgenannten Ansprüche, **dadurch gekennzeichnet**, dass das mindestens eine erste neuronale Netzwerk (10) und das mindestens eine zweite neuronale Netzwerk Teil (11) eines gemeinsamen neuronalen Netzwerks sind.

6. Signalverarbeitungseinrichtung nach einem der vorgenannten Ansprüche, **gekennzeichnet durch** eine Nutzerschnittstelle (18) zum Empfang von Nutzereingaben und/oder zur Ausgabe von Informationen an einen Nutzer.

7. Signalverarbeitungseinrichtung nach einem der vorgenannten Ansprüche, **gekennzeichnet durch** mindestens einen Datenspeicher (15, 17) zum Speichern bekannter Arten von Audiosignalen (a, b1, b2, c, d).

8. System zur Verarbeitung von Audiosignalen aufweisend,
8.1. mindestens eine Signalverarbeitungseinrichtung (4) zur Verarbeitung von Audiosignalen (a, b1, b2, c, d) nach einem der vorgenannten Ansprüche,
8.2. mindestens eine Aufzeichnungseinrichtung (6, 19) zum Aufzeichnen eines Eingangssignals (E), wobei die Aufzeichnungseinrichtung (6, 19) über die Eingangsschnittstelle (13) in datenübertragenderweise mit der Signalverarbeitungseinrichtung (4) verbunden ist, und
8.3. mindestens eine Wiedergabeeinrichtung (7) zum Wiedergeben eines Ausgangssignals (O), wobei die Wiedergabeeinrichtung (7) über die Ausgangsschnittstelle (14) in datenübertragenderweise mit der Signalverarbeitungseinrichtung (4) verbunden ist.

9. System nach Anspruch 8, **dadurch gekennzeichnet**, dass die mindestens eine Signalverarbeitungseinrichtung (4) als mobiles Gerät, insbesondere als Teil eines Mobilfunkgeräts (5), ausgebildet ist.

10. System nach einem der Ansprüche 8 oder 9, **gekennzeichnet durch** einen modularen Aufbau.

11. Verfahren zur Verarbeitung von Audiosignalen mit den Schritten:
11.1. Bereitstellen einer Signalverarbeitungseinrichtung (4) nach einem der Ansprüche 1 bis 7,
11.2. Bereitstellen eines Eingangssignals (E),
11.3. Zuführen des Eingangssignals (E) zur Signalverarbeitungseinrichtung (4) über die Eingangsschnittstelle (13),
11.4. Aufbereiten des Eingangssignals (E) mithilfe des mindestens einen ersten neuronalen Netzwerks (10),
11.5. Separieren eines oder mehrerer Audiosignale (a, b1, b2, c, d) aus dem aufbereiteten Eingangssignal (E) mithilfe des mindesten einen zweiten neuronalen Netzwerks (11),

11.6. Bestimmen eines Prioritätsparameters zu jedem der Audiosignale (a, b1, b2, c, d),
11.7. Modulieren jedes Audiosignals (a, b1, b2, c, d) in Abhängigkeit von dem jeweiligen Prioritätsparameter,
11.8. Kombinieren der Audiosignale (a, b1, b2, c, d) zu einem Ausgangssignal (O),
11.9. Ausgeben des Ausgangssignals (O) über die Ausgangsschnittstelle (14).

12. Verfahren nach Anspruch 11, **dadurch gekennzeichnet**, dass die separierten Audiosignale (a, b1, b2, c, d) klassifiziert werden.

13. Verfahren nach einem der Ansprüche 11 oder 12, **dadurch gekennzeichnet**, dass eine Auswahl des mindestens einen zweiten neuronalen Netzwerks (11) aus einer zur Verfügung stehenden Menge von unterschiedlichen zweiten neuronalen Netzwerken (11) anhand nutzerspezifischer Daten und/oder bereits separierter Audiosignale (a, b1, b2, c, d) erfolgt.

14. Verfahren nach einem der Ansprüche 11 bis 13, **dadurch gekennzeichnet**, dass die Bestimmung der Prioritätsparameter asynchron zu weiteren Schritten des Verfahrens erfolgt.

15. Verfahren nach einem der Ansprüche 11 bis 14, **dadurch gekennzeichnet**, dass die Bestimmung der Prioritätsparameter abhängig von nutzerspezifischen Daten, Vorlieben des Nutzers und/oder einem Informationsgehalt des jeweiligen Audiosignals (a, b1, b2, c, d) erfolgt.

Es folgen 4 Seiten Zeichnungen

Anhängende Zeichnungen

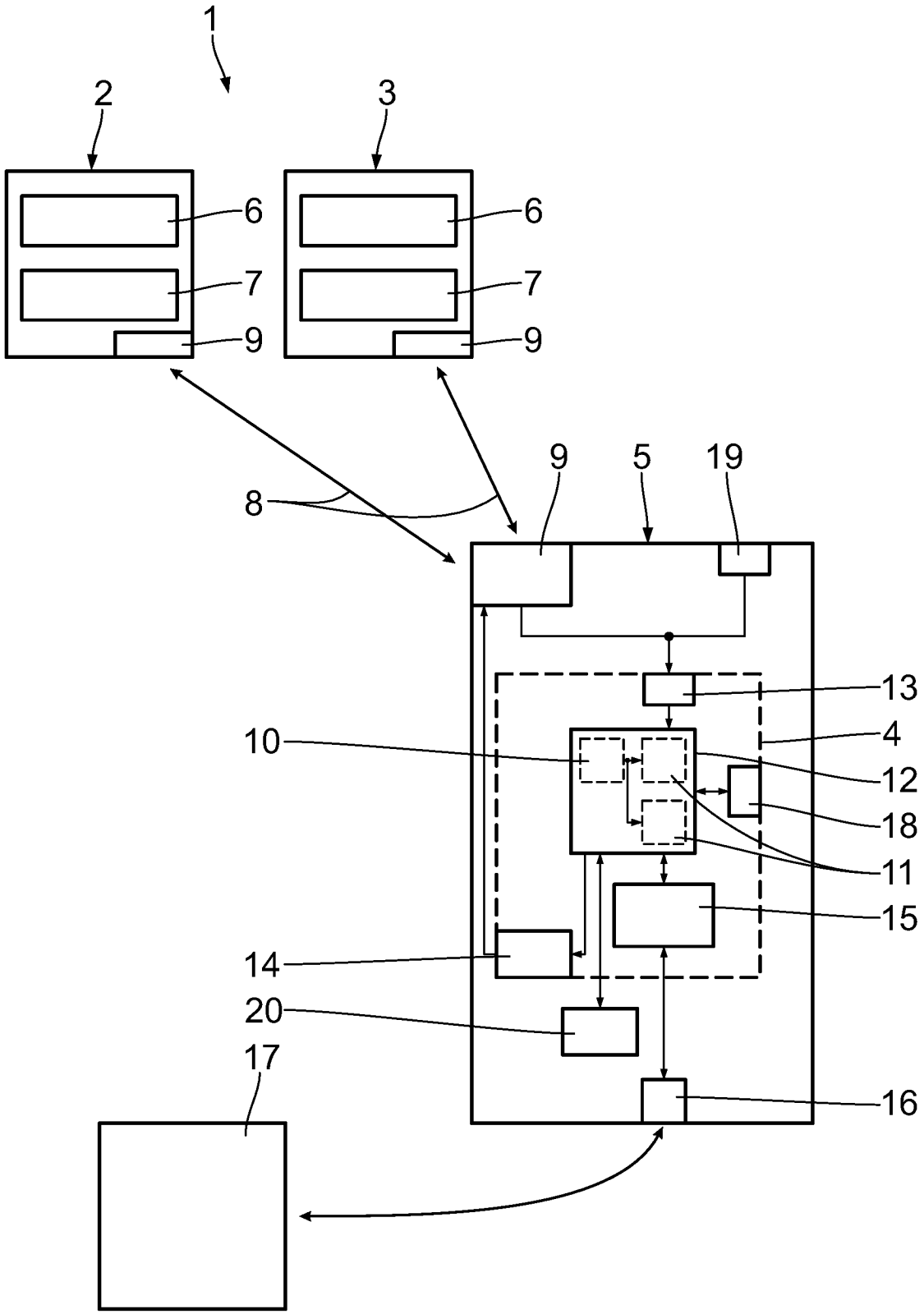


Fig. 1

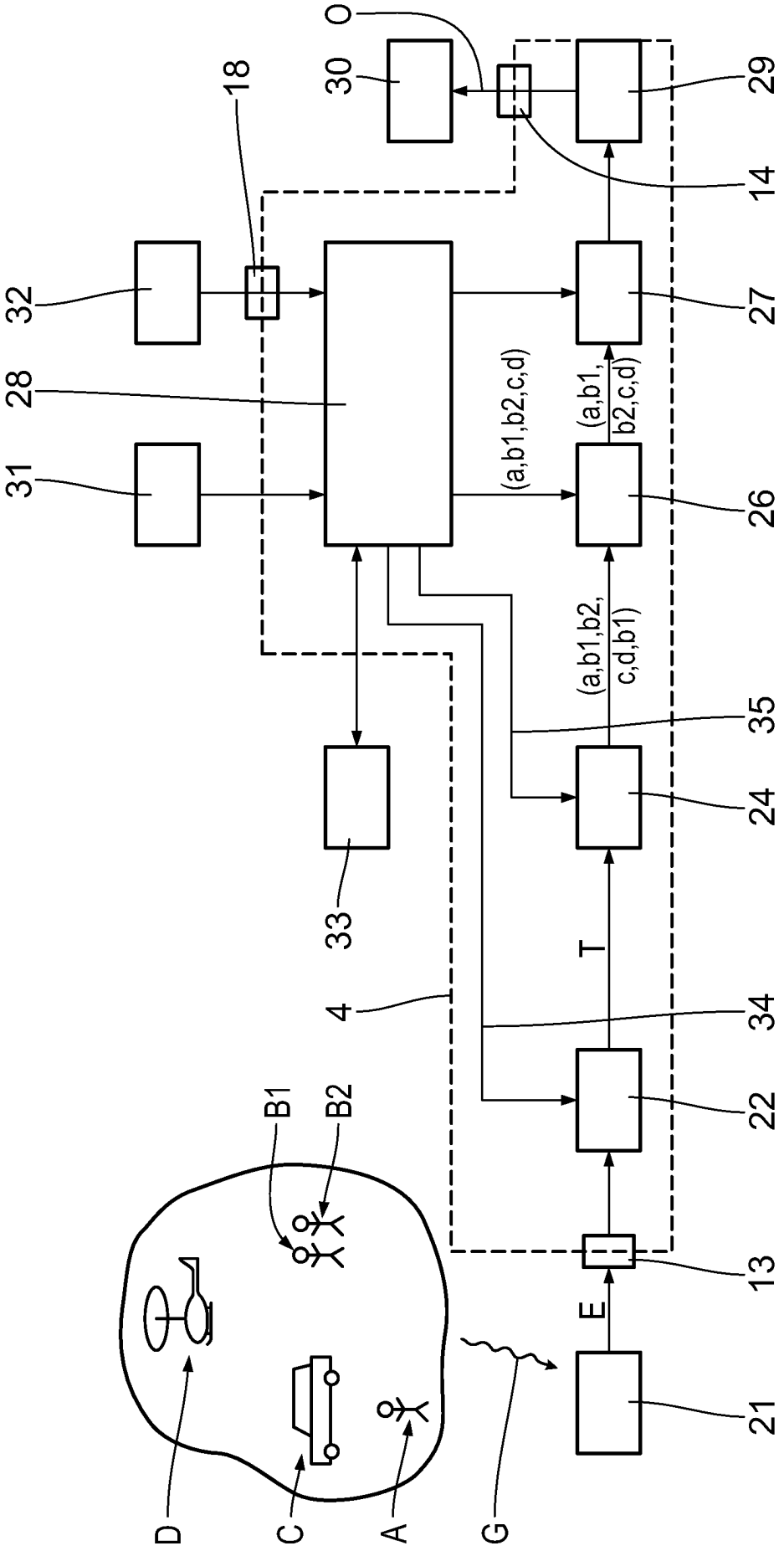

O
30
18
29
14
32
27
28
(a,b1,b2,c,d)
(a,b1,
b2,c,d)
31
26
(a,b1,b2,
c,d,b1)
35
24
33
T
4
34
22
B1
B2
13
E
21
D
C
A
G

Fig. 2

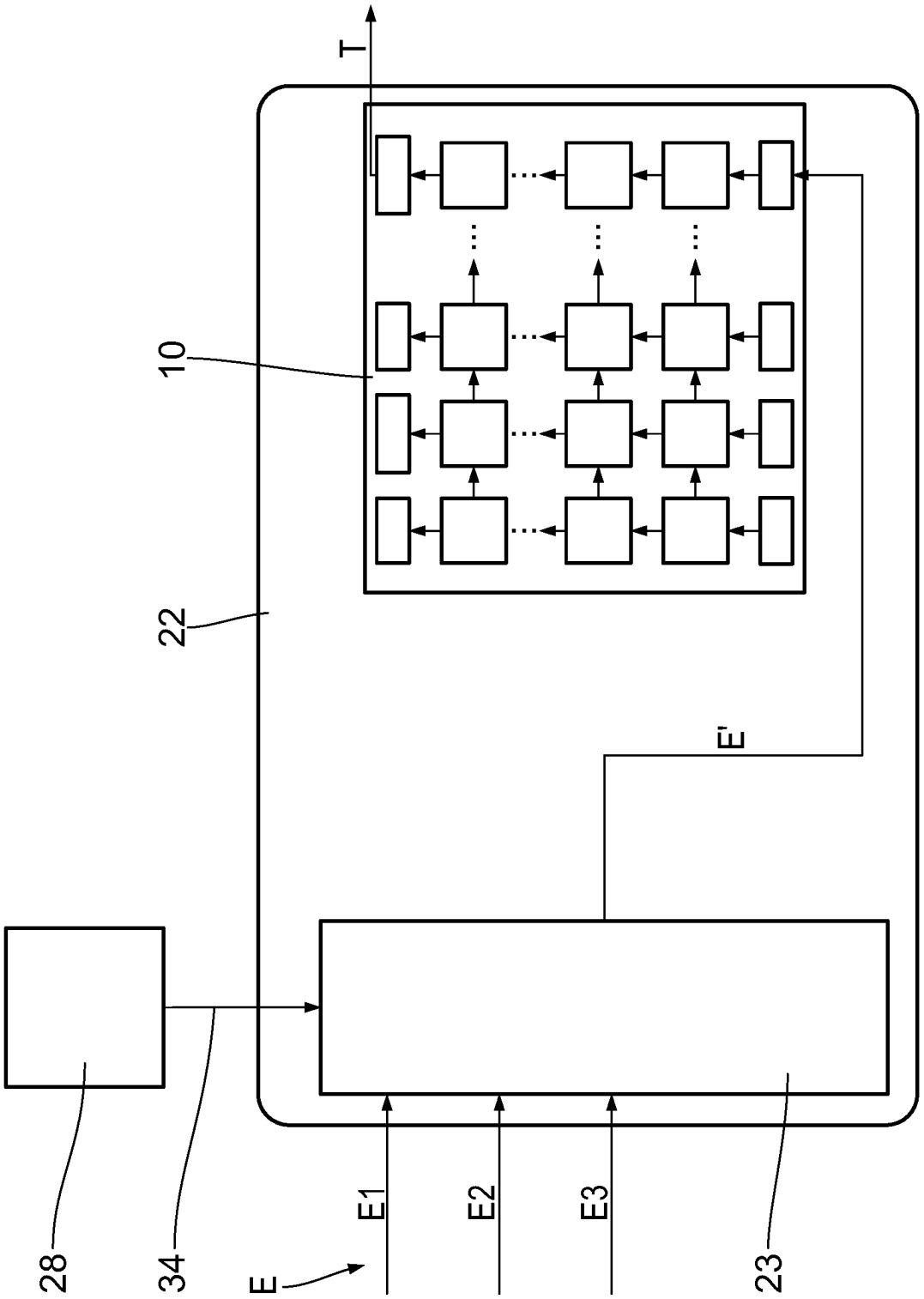


Fig. 3

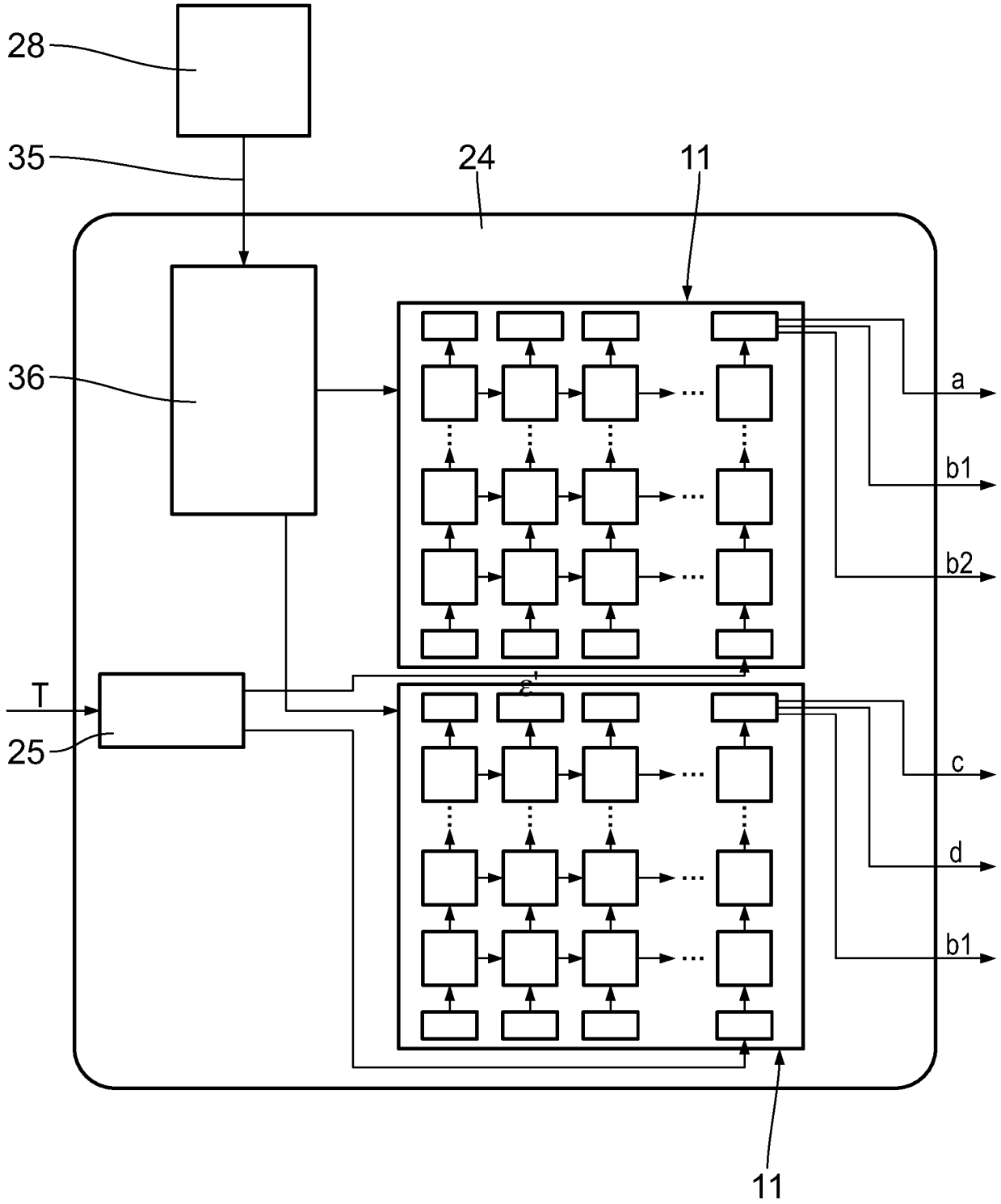


Fig. 4