(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 4 035 154 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
25.06.2025 Patentblatt 2025/26

(21) Anmeldenummer: **21701876.1**

(22) Anmeldetag: **05.02.2021**

(51) Internationale Patentklassifikation (IPC):
G10L 21/0308 *(2013.01)* ***G10L 25/30*** *(2013.01)*

(52) Gemeinsame Patentklassifikation (CPC):
G10L 21/0308; G10L 25/30

(86) Internationale Anmeldenummer:
PCT/EP2021/052776

(87) Internationale Veröffentlichungsnummer:
WO 2021/160533 (19.08.2021 Gazette 2021/33)

(54) **EXTRAKTION EINES AUDIOOBJEKTES**

EXTRACTION OF AN AUDIO OBJECT

EXTRACTION D'UN OBJET AUDIO

(84) Benannte Vertragsstaaten:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priorität: **14.02.2020 DE 102020000974**

(43) Veröffentlichungstag der Anmeldung:
03.08.2022 Patentblatt 2022/31

(73) Patentinhaber: **Lawo Holding AG**
76437 Rastatt (DE)

(72) Erfinder:
- **SCHRÖDER, Leon**
  76437 Rastatt (DE)
- **ZIEGLER, Jonathan**
  76437 Rastatt (DE)

(74) Vertreter: **Lichti - Patentanwälte Partnerschaft mbB**
Postfach 41 07 60
76207 Karlsruhe (DE)

(56) Entgegenhaltungen:
CN-A- 110 534 127

- **LUO YI ET AL: "FaSNet: Low-Latency Adaptive Beamforming for Multi-Microphone Audio Processing", 2019 IEEE AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING WORKSHOP (ASRU), IEEE, 14 December 2019 (2019-12-14), pages 260 - 267, XP033718887, DOI: 10.1109/ASRU46091.2019.9003849**

Anmerkung: Innerhalb von neun Monaten nach Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents im Europäischen Patentblatt kann jedermann nach Maßgabe der Ausführungsordnung beim Europäischen Patentamt gegen dieses Patent Einspruch einlegen. Der Einspruch gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

EP 4 035 154 B1

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Extraktion von mindestens einem Audioobjekt aus mindestens zwei Audio-Eingangssignalen, die jeweils das Audioobjekt enthalten. Ferner betrifft die Erfindung ein System zur Extraktion eines Audioobjektes und ein Computerprogramm mit Programmcodemitteln.

[0002] Im Sinne der Erfindung sind Audioobjekte Audiosignale von Objekten, wie beispielsweise das Geräusch beim Abschießen eines Fußballs, Klatschgeräusche eines Publikums oder der Vortrag eines Gesprächsteilnehmers. Die Extraktion des Audioobjektes im Sinne der Erfindung ist demgemäß die Separation des Audioobjekts von übrigen, störenden Einflüssen, die im Folgenden als Störschall bezeichnet sind. Beispielsweise wird bei der Extraktion eines Schussgeräuschs beim Fußballspiel das reine Schussgeräusch als Audioobjekt von den Geräuschen der Spieler und des Publikums separiert, so dass das Schussgeräusch schließlich als reines Audiosignal vorliegt.

[0003] Aus dem Stand der Technik sind gattungsgemäße Verfahren bekannt, die Extraktion von Audioobjekten vorzunehmen. Eine grundlegende Herausforderung ist dabei, dass üblicherweise die Mikrofone zur Quelle des Audioobjekts unterschiedlich beabstandet sind. Daher befindet sich das Audioobjekt an unterschiedlichen zeitlichen Positionen der Audio-Eingangssignale, was die Auswertung erschwert und verlangsamt.

[0004] Es ist bekannt, die Audio-Eingangssignale derart zu synchronisieren, damit sich das Audioobjekt insbesondere an der jeweils gleichen zeitlichen Position der Audio-Eingangssignale befindet. Dies wird üblicherweise auch als Laufzeitkompensation bezeichnet. Übliche Verfahren nutzen diesbezüglich neuronale Netzwerke. Dabei ist es erforderlich, dass das neuronale Netzwerk auf sämtliche mögliche Mikrofonabstände zur Quelle des Audioobjektes trainiert werden muss. Gerade bei dynamischen Audioobjekten, wie im Falle von Sportveranstaltungen, ist ein effektives Training des neuronalen Netzes aber nicht durchführbar.

[0005] Ferner sind gattungsgemäße Verfahren bekannt, bei denen zur Synchronisierung der Audio-Eingangssignale deren Korrelation, beispielsweise deren Kreuzkorrelation, analytisch berechnet wird, was zwar die Geschwindigkeit des Verfahrens steigert, aber die Zuverlässigkeit der nachfolgenden Extraktion des Audioobjekts beeinträchtigt, da die Korrelation stets unabhängig von der Art des Audioobjekts berechnet wird. Dabei werden aber oft für die nachfolgende Extraktion des Audioobjekts störende Effekte, insbesondere Störschall, verstärkt.

[0006] Bekannte gattungsgemäße Verfahren werden in CN 110 534 127 A sowie in Luo et al: "FaSNet: Low-Latency Adaptive Beamforming for Multi-Microphone Audio Processing", AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING WORKSHOP (ASRU), IEEE, 14. Dezember 2019, Seiten 260-267, XP033718887, DOI: 10.1109/ASRU46091.2019.9003849 [gefunden am 2020-02-19] beschrieben.

[0007] Es ist daher die Aufgabe der Erfindung, die genannten Nachteile aus dem Stand der Technik zu beseitigen und insbesondere die Zuverlässigkeit der Extraktion des Audioobjektes zu verbessern bei gleichzeitiger Optimierung der Geschwindigkeit des Verfahrens.

[0008] Die Aufgabe wird gelöst durch ein Verfahren mit den Merkmalen des Anspruchs 1, ein System mit den Merkmalen des Anspruchs 11 und ein Computerprogramm mit den Merkmalen des Anspruchs 14. Bevorzugte Ausführungsformen der Erfindung werden in den abhängigen Ansprüchen 2-10, 12 und 13 dargelegt.

[0009] Die Erfindung basiert auf der Grundüberlegung, dass durch die analytische Berechnung der Korrelation, beispielsweise der Kreuzkorrelation, die Qualität des extrahierten Audioobjekts, also die Signaltrennungsqualität des Verfahrens, verbessert wird. Gleichwohl wird durch die Ausbildung des ersten und des zweiten trainierten Operators eine Möglichkeit geschaffen, mit Hilfe von trainierten Komponenten die Zuverlässigkeit der nachfolgenden Extraktion des Audioobjektes zu verbessern. Insofern stellt die Erfindung ein neuartiges Verfahren dar, das die Extraktion des Audioobjektes zuverlässig und schnell durchführt. Dadurch ist das Verfahren auch bei komplexen Mikrofongeometrien, wie beispielsweise großen Mikrofonabständen einsetzbar.

[0010] Der erste trainierte Operator kann eine insbesondere trainierte Transformation der Audio-Eingangssignale in einen Merkmalsraum umfassen, um die nachfolgenden Verfahrensschritte zu vereinfachen. Der zweite trainierte Operator kann mindestens eine Normierung des Korrelationsvektors umfassen, um die Genauigkeit der Berechnung des synchronisierten zweiten Audio-Eingangssignals zu verbessern. Ferner kann der zweite trainierte Operator eine zur Transformation des ersten trainierten Operators inverse Transformation des synchronisierten zweiten Audio-Eingangssignals, insbesondere zurück in den Zeitraum der Audio-Eingangssignale, vorsehen.

[0011] Erfindungsgemäß weist der zweite trainierte Operator ein iteratives Verfahren mit endlich vielen Iterationsschritten auf, wobei insbesondere in jedem Iterationsschritt ein Synchronisationsvektor, vorzugsweise ein optimierter Korrelationsvektor, insbesondere ein optimierter Kreuzkorrelationsvektor, bestimmt werden, was eine Beschleunigung des erfindungsgemäßen Verfahrens bewirkt. Die Anzahl der Iterationsschritte des zweiten trainierten Operators kann benutzerseitig definierbar sein, um das Verfahren benutzerseitig zu konfigurieren.

[0012] In jedem Iterationsschritt des zweiten trainierten Operators erfolgt vorzugsweise eine gestreckte Faltung des Audio-Signals mit mindestens einem Teil des Synchronisationsvektors, insbesondere des optimierten Korrelationsvektors. In jedem Iterationsschritt kann eine Normierung des Synchronisationsvektors und/oder eine gestreckte Faltung des synchronisierten Audio-Eingangssignals mit dem Synchronisationsvektor erfolgen, um die Signaltrennungsqualität

des Verfahrens zu verbessern.

[0013] In einer weiteren Ausgestaltung der Erfindung sieht der zweite trainierte Operator die Bestimmung mindestens einer akustischen Modellfunktion vor. Im Sinne der Erfindung entspricht die akustische Modellfunktion insbesondere dem Zusammenhang zwischen dem Audioobjekt und dem aufgenommenen Audio-Eingangssignal. Damit gibt die akustische Modellfunktion beispielsweise die akustischen Eigenschaften der Umgebung, wie etwa akustische Reflexionen (Hall), frequenzabhängige Absorptionen und/oder Bandpass-Effekte wieder. Außerdem beinhaltet die akustische Modellfunktion insbesondere die Aufnahmecharakteristik mindestens eines Mikrofons. Insofern ist durch den zweiten trainierten Operator im Rahmen der Optimierung des Korrelationsvektors die Kompensation unerwünschter akustischer Effekte auf das Audiosignal, bedingt etwa durch die Umgebung und/oder die Aufnahmecharakteristik des mindestens einen Mikrofons möglich. Neben der Kompensation der Laufzeit ist damit auch die Kompensation störender akustischer Einflüsse, beispielsweise bedingt durch den Propagationsweg des Schalls, möglich, was die Signaltrennungsqualität des erfindungsgemäßen Verfahrens verbessert.

[0014] Das trainierte Modell zum Extrahieren des Audioobjektes kann mindestens eine Transformation des ersten Audio-Eingangssignals und des synchronisierten zweiten Audio-Eingangssignals jeweils in einen insbesondere höherdimensionalen Darstellungsraum vorsehen, was die Signaltrennungsqualität verbessert. Im Sinne der Erfindung weist der Darstellungsraum eine im Vergleich zu dem in der Regel eindimensionalen Zeitraum der Audio-Eingangssignale höhere Dimensionalität auf. Indem die Transformationen als Teile eines neuronalen Netzwerks ausgebildet sein können, können die Transformationen spezifisch hinsichtlich des zu extrahierenden Audioobjektes trainiert sein.

[0015] Das trainierte Modell des Extrahierens des Audioobjekts kann die Anwendung mindestens einer trainierten Filtermaske auf das erste Audio-Eingangssignal und auf das synchronisierte zweite Audio-Eingangssignal vorsehen. Die trainierte Filtermaske ist vorzugsweise spezifisch auf das Audioobjekt trainiert.

[0016] Das trainierte Modell des Extrahierens des Audioobjekts kann mindestens eine Transformation des Audioobjekts in den Zeitraum der Audio-Eingangssignale vorsehen, um insbesondere eine vorausgegangene Transformation in den Darstellungsraum rückgängig zu machen.

[0017] Die Verfahrensschritte des Synchronisierens und/oder des Extrahierens und/oder der Ausgabe des Audioobjektes sind vorzugsweise einem einzigen neuronalen Netzwerk zugeordnet, um ein spezifisches Training des neuronalen Netzwerks hinsichtlich des Audioobjektes zu ermöglichen. Durch die Ausgestaltung eines einzigen neuronalen Netzwerks wird die Zuverlässigkeit des Verfahrens und dessen Signaltrennungsqualität insgesamt verbessert.

[0018] Vorzugsweise wird das neuronale Netzwerk mit Soll-Trainingsdaten trainiert, wobei die Soll-Trainingsdaten Audio-Eingangssignale und dazu korrespondierende vordefinierte Audioobjekte umfassen, mit den folgenden Trainingsschritten: Vorwärtsspeisen des neuronalen Netzwerks mit den Soll-Trainingsdaten unter Erhalt eines ermittelten Audioobjekts, Bestimmen eines Fehlerparameters, insbesondere eines Fehlervektors zwischen dem ermittelten Audioobjekt und dem vordefinierten Audioobjekt und Ändern von Parametern des neuronalen Netzwerks durch Rückwärtsspeisen des neuronalen Netzwerks mit dem Fehlerparameter, insbesondere mit dem Fehlervektor, falls ein Qualitätsparameter des Fehlerparameters, insbesondere des Fehlervektors, einen vordefinierten Wert übersteigt.

[0019] Das Training ist dabei auf das spezifische Audioobjekt ausgerichtet; mindestens zwei Parameter der trainierten Komponenten des erfindungsgemäßen Verfahrens können wechselseitig voneinander abhängig sein.

[0020] Vorzugsweise ist das Verfahren derart ausgestaltet, dass es kontinuierlich abläuft, was auch als "Online-Betrieb" bezeichnet ist. Im Sinne der Erfindung werden dabei ständig, insbesondere ohne Benutzereingabe, Audio-Eingangssignale eingelesen und zur Extraktion von Audioobjekten ausgewertet. Dabei können beispielsweise die Audio-Eingangssignale jeweils Teile von insbesondere kontinuierlich eingelesenen Audio-Signalen mit insbesondere vordefinierter Länge sein. Dies wird auch als "Buffering" bezeichnet. Besonders vorzugsweise kann das Verfahren derart ausgebildet sein, dass die Latenz des Verfahrens höchstens 100 ms, insbesondere höchstens 80 ms, vorzugsweise höchstens 40 ms beträgt. Latenz ist im Sinne der Erfindung die Laufzeit des Verfahrens, gemessen ab dem Einlesen der Audio-Eingangssignale bis zur Ausgabe des Audioobjektes. Ein Betrieb des Verfahrens ist daher in Echtzeit möglich.

[0021] Das erfindungsgemäße System kann ein erstes Mikrofon zum Empfangen des ersten Audio-Eingangssignals und ein zweites Mikrofon zum Empfangen des zweiten Audio-Eingangssignals vorsehen, wobei die Mikrofone jeweils mit dem System derart verbindbar sind, dass die Audio-Eingangssignale der Mikrofone der Steuereinheit des Systems zuführbar sind. Das System kann insbesondere als Komponente eines Mischpults ausgestaltet sein, mit dem die Mikrofone verbindbar sind. Besonders vorzugsweise ist das System ein Mischpult. Die Verbindung des Systems mit dem Mikrofonen kann kabelgebunden und/oder kabellos sein. Das Computerprogramm zur Durchführung des erfindungsgemäßen Verfahrens ist vorzugsweise auf einer Steuereinheit des erfindungsgemäßen Systems ausführbar.

[0022] Weitere Vorteile und Merkmale der Erfindung ergeben sich aus den Ansprüchen und der nachfolgenden Beschreibung, in der Ausgestaltungen der Erfindung unter Bezugnahme auf die Zeichnungen im Einzelnen erläutert sind. Dabei zeigen:

Fig. 1 Ein erfindungsgemäßes System in einer schematischen Ansicht;

Fig. 2 eine Übersicht eines erfindungsgemäßen Verfahrens in einem Ablaufdiagramm mit modellhaften Signalen;

Fig. 3 ein Ablaufdiagramm zum Verfahrensschritt einer Synchronisierung von Audio-Eingangssignalen mit modellhaften Signalen;

Fig. 4 ein Ablaufdiagramm zu einem iterativen Verfahren der Synchronisierung;

Fig. 5 ein Ablaufdiagramm zum Extrahieren des Audioobjektes und

Fig. 6 ein Ablaufdiagramm zum Trainieren des erfindungsgemäßen Verfahrens.

[0023] Fig. 1 zeigt eine Ausgestaltung eines erfindungsgemäßen Systems 10 zur Extraktion eines Audioobjektes 11 in einer schematischen Darstellung, wobei das System 10 ein Mischpult 10a ist. Audioobjekte 11 im Sinne der Erfindung sind akustische Signale, die einem Ereignis und/oder einem Objekt zugeordnet sind. Im vorliegenden Ausführungsbeispiel der Erfindung ist das Audioobjekt 11 das Geräusch 12 eines abgeschossenen, in Fig. 1 nicht dargestellten Fußballs.

[0024] Das Geräusch 12 wird von zwei Mikrofonen 13, 14 aufgenommen, die jeweils ein Audio-Eingangssignal a1, a2 erzeugen, so dass die Audio-Eingangssignale a1, a2 das Geräusch 12 enthalten. Aufgrund der unterschiedlichen Distanzen der Mikrofone 13, 14 zum Geräusch 12 befindet sich das Geräusch 12 an unterschiedlichen zeitlichen Positionen der Audio-Eingangssignale a1, a2. Zusätzlich unterscheiden sich die Audio-Eingangssignale a1, a2 aufgrund der akustischen Eigenschaften der Umgebung voneinander und weisen daher jeweils auch unerwünschte Anteile auf, die beispielsweise durch die Propagationsstrecken des Schalls bis zu den Mikrofonen 13, 14 etwa in Form von Hall und/oder unterdrückten Frequenzen, verursacht sind, und die im Sinne der Erfindung als Störschall bezeichnet werden. Im Sinne der Erfindung gibt eine erste akustische Modellfunktion M1 die akustischen Einflüsse der Umgebung und der Aufnahmecharakteristik des Mikrofons 13 auf das aufgenommene Audio-Eingangssignal a1 des ersten Mikrofons 13 wieder. Das Audio-Eingangssignal a1 entspricht mathematisch insofern einer Faltung des Geräuschs 12 mit der ersten akustischen Modellfunktion M1. Analog gilt dies für eine zweite akustische Modellfunktion M2 und für das aufgenommene Audio-Eingangssignal a2 des zweiten Mikrofons 14.

[0025] Die Mikrofone 13, 14 sind mit dem Mischpult 10a verbunden, so dass die Audio-Eingangssignale a1, a2 an eine Steuereinheit 15 des Systems 10 übermittelt werden, damit die Steuereinheit 15 die Audio-Eingangssignale a1, a2 auswertet und das Geräusch 12 aus den Audio-Eingangssignalen a1, a2 mit Hilfe des erfindungsgemäßen Verfahrens extrahiert und zur weiteren Verwendung ausgibt. Bei der Steuereinheit 15 zur Extraktion des Audioobjektes 11 handelt es sich um einen Mikrokontroller und/oder um einen Programmcodeblock eines entsprechenden Computerprogramms. Die Steuereinheit 15 umfasst ein trainiertes neuronales Netzwerk, das mit Audio-Eingangssignalen a1, a2 insbesondere vorwärts gespeist wird. Das neuronale Netzwerk ist dazu trainiert, das spezifische Audioobjekt 11, also im vorliegenden Falle das Geräusch 12, aus den Audio-Eingangssignalen a1, a2 zu extrahieren und insbesondere von Störschall-Anteilen der Audio-Eingangssignale a1, a2 zu trennen. Im Wesentlichen werden dabei die Auswirkungen der akustischen Modellfunktionen M1, M2 auf das Geräusch 12 in den Audio-Eingangssignalen a1, a2 kompensiert.

[0026] Fig. 2 veranschaulicht eine Ausgestaltung des erfindungsgemäßen Verfahrens in einer Übersicht als Flussdiagramm mit modellhaften Audio-Eingangssignalen a1, a2, an denen das Verfahren durchgeführt wird. In einem ersten Schritt V1 erfolgt ein Synchronisieren des zweiten Audio-Eingangssignals a2 mit dem ersten Audio-Eingangssignal a1, so dass im Ergebnis ein synchronisiertes zweites Audio-Eingangssignal a2' erhalten wird. Im Sinne der Erfindung weist das synchronisierte zweite Audio-Eingangssignal a2' insbesondere das Geräusch 12 an im Wesentlichen der gleichen zeitlichen Position auf wie das erste Audio-Eingangssignal a1, was die nachfolgenden Verfahrensschritte maßgeblich beschleunigt und vereinfacht. Insofern entspricht die Synchronisierung V1 der Audio-Eingangssignale a1, a2 insbesondere einer Kompensation der Laufzeitdifferenzen zwischen den Audio-Eingangssignalen a1, a2.

[0027] Anschließend erfolgt gemäß Fig. 2 das Extrahieren V2 des Geräuschs 12 durch die Anwendung eines trainierten Modells auf das erste Audio-Eingangssignal a1 und auf das synchronisierte zweite Audio-Eingangssignal a2', so dass im Ergebnis das Geräusch 12 als Audiosignal erhalten wird. Das trainierte Modell ist dem neuronalen Netzwerk zugeordnet und ist als ein Teil von diesem auf die Extraktion des spezifischen Audioobjekts 11, hier des Geräuschs 12, trainiert. Im nachfolgenden Verfahrensschritt erfolgt die Ausgabe V3 des Geräuschs 12 als Audio-Ausgangssignal Z.

[0028] Die Verfahrensschritte des Synchronisierens V1, des Extrahierens V2 des Geräuschs 12 und dessen Ausgabe V3 sind einem einzigen, trainierten neuronalen Netzwerk zugeordnet, so dass das Verfahren als End-to-End-Verfahren ausgebildet ist. Dadurch ist es als Ganzes trainiert und läuft automatisch und kontinuierlich ab, wobei die Extraktion des Geräuschs in Echtzeit, also mit einer Latenz von höchstens 40 ms erfolgt.

[0029] Fig. 3 zeigt einen Verfahrensablauf des Synchronisierens V1 der Audio-Eingangssignale a1, a2 in einem Flussdiagramm mit modellhaften Audio-Eingangssignalen a1, a2 zur Veranschaulichung der Verfahrensschritte. In einem ersten Verfahrensschritt V4 der Fig. 3 wird ein erster trainierter Operator des neuronalen Netzwerks jeweils auf die Audio-Eingangssignale a1, a2 angewendet, um Audio-Signale m1, m2 zu generieren. In einer Ausgestaltung der Erfindung

werden die Audio-Eingangssignale a1, a2 durch den ersten trainierten Operator des neuronalen Netzwerks in einen im Vergleich zu den Audio-Eingangssignalen a1, a2 höherdimensionalen Merkmalsraum in der Zeitdomäne zu den Audio-Signalen m1, m2 transformiert, um die nachfolgenden Berechnungen zu vereinfachen und zu beschleunigen. Je nach Art des Audioobjekts 11 erfolgt bereits bei der Transformation eine Bearbeitung der Audio-Signale m1, m2. Die transformierten Audio-Signale m1, m2 sind in Fig. 3 modellhaft dargestellt.

[0030] Im zweiten Verfahrensschritt V5 der Fig. 3 erfolgt die analytische Berechnung der Kreuzkorrelation als Korrelation zwischen den Audio-Signalen m1, m2, die mathematisch wie folgt definiert ist:

$$(m_1 \star m_2)[t] \quad \hat{=} \sum_{n=-\infty}^{\infty} m_1[n]\, m_2[n+t]$$

[0031] Die Berechnung V5 resultiert in einen Kreuzkorrelationsvektor k, der modellhaft in Fig. 3 dargestellt ist. Im dritten Verfahrensschritt V6 wird der Kreuzkorrelationsvektor k mit Hilfe eines zweiten trainierten Operators des neuronalen Netzwerks optimiert, wobei mittels des zweiten trainierten Operators die Berechnung der akustischen Modellfunktion M erfolgt, um deren Auswirkungen auf die Audio-Signale m1, m2 zu kompensieren. Der zweite trainierte Operator dient damit beispielsweise als akustischer Filter und sieht im Ausführungsbeispiel der Fig. 3 insbesondere eine Normierung des Kreuzkorrelationsvektors k vor, beispielsweise mittels einer Softmax-Funktion. Der dadurch erhaltene Synchronisationsvektor s ist modellhaft in Fig. 3 dargestellt.

[0032] Im vierten Verfahrensschritt der Fig. 3 erfolgt die Berechnung V7 des synchronisierten zweiten Audio-Eingangssignals a2' durch die Faltung des Synchronisationsvektors s mit dem zweiten Audio-Eingangssignal a2.

[0033] Das synchronisierte zweite Audio-Eingangssignal a2' ist in Fig. 3 modellhaft dargestellt. Im Vergleich zum ursprünglichen Audio-Eingangssignal a2 ist erkennbar, dass im hier betrachteten, stark vereinfachten Modell eine Kompensation der Laufzeitdifferenz als zeitlicher Offset erfolgt ist. Das synchronisierte zweite Audio-Eingangssignal a2' wird anschließend, wie bereits beschrieben, für die Extraktion V2 des Audioobjekts 11 verwendet.

[0034] Fig. 4 zeigt eine weitere Ausgestaltung der Synchronisierung V1 der Audio-Eingangssignale a1, a2, bei der ein iteratives Verfahren zur Beschleunigung der Berechnung vorgesehen ist, wobei die Anzahl der Iterationsschritte I benutzerseitig festgelegt ist. Im ersten Iterationsschritt erfolgt eine Berechnung des Korrelationsvektors zwischen den Audio-Signalen m1, m2 ähnlich dem Verfahren gemäß Fig. 3 bis zur Berechnung V7 des synchronisierten Audio-Eingangssignals a2', wobei der Synchronisationsvektor s_i des aktuellen Iterationsschritts i aber nun im Rahmen der Optimierung V6 bei jedem Iterationsschritt i mittels der maxpool-Funktion beschränkt wird. Anschließend erfolgt - in jedem Iterationsschritt i - die Berechnung V8 des iterativen Audio-Signals $m2_i$ für die Iterationsstufe i mittels einer gestreckten Faltung, die mathematisch wie folgt definiert ist:

$$\left(a_2 *_{d_i} s\right)(t) = \sum_{n=-d_i}^{d_i} a_2(d_i \cdot n)\, s\,(n+t)$$

[0035] Der Faktor d_i entspricht dabei dem Maß der Beschränkung des Kreuzkorrelationsvektors für den Iterationsschritt i, wobei die Summierung über den Faktor d_i erfolgt. Dieser Vorgang wird so lange wiederholt, bis die benutzerseitig vorgegebene Anzahl an Iterationsschritten I durchgeführt wurde. Schließlich erfolgt eine gestreckte Faltung V9 des Audio-Signals m2 mit dem zuletzt berechneten Synchronisationsvektor S_i, woraufhin das synchronisierte zweite AudioSignal a2'berechnet und ausgegeben wird V7. Durch die Berechnung des Synchronisationsvektors s auf der Basis des Teilbereichs der im vorigen Iterationsschritt ermittelten Parameter reduziert sich die Komplexität der Berechnungen, was die Laufzeit des Verfahrens beschleunigt, ohne dessen Genauigkeit zu beeinträchtigen.

[0036] Fig. 5 zeigt eine Ausgestaltung der Extraktion V2 des Audioobjektes 11 aus dem Audio-Eingangssignal a1 und dem synchronisierten zweiten Audio-Eingangssignal a2' in einem Flussdiagramm. In einem ersten Verfahrensschritt V10 werden die Audio-Eingangssignale a1, a2' durch die Anwendung eines ersten trainierten Modells des neuralen Netzwerks jeweils in einen höherdimensionalen Darstellungsraum transformiert, um die nachfolgenden Berechnungen zu vereinfachen. Beispielsweise weist das erste trainierte Modell eine gängige Filterbank mit insbesondere einer Terzbandfilterbank und/oder einer Mel-Filterbank auf, wobei die Parameter der Filter durch das vorausgegangene Training des neuronalen Netzwerks optimiert worden sind.

[0037] Im zweiten Verfahrensschritt V11 erfolgt die Separation des Audioobjekts 11 von den Audio-Eingangssignalen a1, a2' durch Anwendung eines zweiten trainierten Modells des neuronalen Netzwerks auf die Audio-Eingangssignale a1, a2'. Auch die Parameter des zweiten trainierten Modells wurden durch das vorausgegangene Training optimiert und sind

insbesondere von dem ersten trainierten Modell des vorangehenden Verfahrensschrittes V10 abhängig. Im Ergebnis dieses Verfahrensschrittes V11 wird das Audioobjekt 11 aus den Audio-Eingangssignalen a1, a2' erhalten und befindet sich noch im höherdimensionalen Darstellungsraum.

[0038] Im dritten Verfahrensschritt V12 der Fig. 5 wird das separierte Audioobjekt 11 durch die Anwendung eines dritten trainierten Modells des neuronalen Netzwerks auf das Audioobjekt 11 in den ursprünglichen, eindimensionalen Zeitraum der Audiosignale a1, a2 transformiert, wobei die Parameter des dritten trainierten Modells von jenen der übrigen trainierten Modelle abhängig sind und durch das vorausgegangene Training gemeinsam optimiert wurden. Insofern ist das dritte trainierte Modell der Transformation gemäß dem dritten Verfahrensschritt V12 der Fig. 5 funktional als Komplement zur Transformation V10 gemäß dem ersten trainierten Modell zu sehen. Falls beispielsweise im ersten trainierten Modell des ersten Verfahrensschrittes V10 eine eindimensionale Faltung vorgesehen ist, erfolgt in der Rücktransformation V12 eine transponierte eindimensionale Faltung.

[0039] Damit das neuronale Netzwerk das Audioobjekt 11 zuverlässig aus den Audio-Eingangssignalen a1, a2 extrahieren kann, muss es vor dem Einsatz trainiert werden. Dies geschieht beispielsweise durch die nachfolgend beschriebenen Trainingsschritte V13 bis V19, die in Fig. 6 in einem schematischen Ablaufdiagramm gezeigt sind. In den betrachteten Ausführungsbeispielen des erfindungsgemäßen Verfahrens sind die genannten Verfahrensschritte einem einzigen neuronalen Netzwerk zugeordnet und jeweils differenzierbar, so dass mit dem nachfolgend beschriebenen Trainingsverfahren V13 sämtliche trainierten Komponenten spezifisch hinsichtlich des Audioobjekts 11 trainiert werden.

[0040] Vordefinierte Audioobjekte 16 werden mittels vordefinierter Algorithmen zu vorgegebenen Audio-Eingangssignalen a1, a2 generiert V14. Die vordefinierten Audioobjekte 16 sind stets vom gleichen Typ, so dass das Verfahren spezifisch hinsichtlich eines Typs von Audioobjekten 16 trainiert wird. Die generierten Audio-Eingangssignale a1, a2 durchlaufen das erfindungsgemäße Verfahren gemäß Fig. 2 und werden dabei insbesondere durch das neurale Netzwerk vorwärts gespeist V15. Das dadurch ermittelte Audioobjekt 17 wird mit dem vordefinierten Audioobjekt 16 verglichen, um auf dieser Grundlage einen mathematischen Fehlervektor P zu bestimmen V16. Danach erfolgt eine Abfrage V17, ob ein Qualitätsparameter des Fehlervektors P einen vordefinierten Wert unterschreitet und das ermittelte Audioobjekt 17 hinreichend gut extrahiert wurde.

[0041] Überschreitet der Qualitätsparameter den vordefinierten Wert, ist das Abbruchkriterium nicht erfüllt und es wird im nächsten Verfahrensschritt V18 der Gradient des Fehlervektors P bestimmt und rückwärts durch das neuronale Netzwerk gespeist, so dass sämtliche Parameter des neuronalen Netzwerks angepasst werden. Anschließend wird das Trainingsverfahren V13 mit weiteren Datensätzen solange wiederholt, bis der Fehlervektor P einen hinreichend guten Wert erreicht und die Abfrage V17 ergibt, dass das Abbruchkriterium erfüllt wurde. Dann wird der Trainingsprozess V13 abgeschlossen V19 und das Verfahren kann auf reale Daten angewendet werden. Idealerweise werden als vordefinierte Audioobjekte 16 in der Trainingsphase jene Audioobjekte 11 verwendet, die in der Anwendung des Verfahrens auch ermittelt werden sollen, beispielsweise bereits aufgezeichnete Schussgeräusche 12 von Fußbällen.

Patentansprüche

1. Verfahren zur Extraktion von mindestens einem Audioobjekt (11) aus mindestens zwei Audio-Eingangssignalen (a1, a2), die jeweils das Audioobjekt (11) enthalten, mit den folgenden Schritten:

   - Synchronisieren (V1) des zweiten Audio-Eingangssignals (a2) mit dem ersten Audio-Eingangssignal (a1) unter Erhalt eines synchronisierten zweiten Audio-Eingangssignals (a2'),
   - Extrahieren (V2) des Audioobjekts (11) durch die Anwendung von mindestens einem trainierten Modell auf das erste Audio-Eingangssignal (a1) und auf das synchronisierte zweite Audio-Eingangssignal (a2') und
   - Ausgabe (V3) des Audioobjekts (11),

   wobei der Verfahrensschritt des Synchronisierens (V1) des zweiten Audio-Eingangssignals (a2) mit dem ersten Audio-Eingangssignal (a1) die folgenden Verfahrensschritte umfasst:

   - Generieren (V4) von Audio-Signalen (m1, m2) durch Anwendung eines ersten trainierten Operators auf die Audio-Eingangssignale (a1, a2),
   - Analytische Berechnung (V5) einer Korrelation zwischen den Audio-Signalen (m1, m2) unter Erhalt eines Korrelationsvektors (k),
   - Optimieren (V6) des Korrelationsvektors (k) mit Hilfe eines zweiten trainierten Operators unter Erhalt eines Synchronisationsvektors (s) und
   - Bestimmen (V7) des synchronisierten zweiten Audio-Eingangssignals (a2') mit Hilfe des Synchronisationsvektors (s),

dadurch gekennzeichnet, dass der zweite trainierte Operator ein iteratives Verfahren mit endlich vielen Iterationsschritten (I) aufweist.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** der erste trainierte Operator eine insbesondere trainierte Transformation der Audio-Eingangssignale (a1, a2) in einen Merkmalsraum umfasst.

3. Verfahren nach einem der Ansprüche 1 oder 2, **dadurch gekennzeichnet, dass** der zweite trainierte Operator mindestens eine Normierung des Korrelationsvektors (k) umfasst.

4. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** in jedem Iterationsschritt ein Synchronisationsvektor (s) bestimmt wird.

5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet, dass** die Anzahl der Iterationsschritte (I) des zweiten trainierten Operators benutzerseitig definierbar ist.

6. Verfahren nach einem der Ansprüche 4 bis 5, **dadurch gekennzeichnet, dass** in jedem Iterationsschritt eine Normierung des Synchronisationsvektors (s) und/oder eine gestreckte Faltung des synchronisierten Audio-Eingangssignals (a2') mit Synchronisationsvektor (s') erfolgt.

7. Verfahren nach einem der Ansprüche 1 bis 6, **dadurch gekennzeichnet, dass** der zweite trainierte Operator die Bestimmung mindestens einer akustischen Modellfunktion (M) vorsieht.

8. Verfahren nach einem der Ansprüche 1 bis 7, **dadurch gekennzeichnet, dass** das trainierte Modell des Extrahierens (V2) des Audioobjekts (11) mindestens eine Transformation des ersten Audio-Eingangssignals (a1) und des synchronisierten zweiten Audio-Eingangssignals (a2') jeweils in einen insbesondere höherdimensionalen Darstellungsraum vorsieht und/oder dass das trainierte Modell des Extrahierens (V2) des Audioobjekts (11) die Anwendung mindestens einer gelernten Filtermaske auf das erste Audio-Eingangssignal (a1) und auf das synchronisierte zweite Audio-Eingangssignal (a2') vorsieht, wobei insbesondere vorgesehen ist, dass das trainierte Modell des Extrahierens (V2) des Audioobjekts (11) mindestens eine Transformation des Audioobjekts (11) in den Zeitraum der Audio-Eingangssignale (a1, a2) vorsieht.

9. Verfahren nach einem der Ansprüche 1 bis 8, **dadurch gekennzeichnet, dass** die Verfahrensschritte des Synchronisierens (V1) und/oder des Extrahierens (V2) und/oder der Ausgabe (V3) des Audioobjekts (11) einem einzigen neuronalen Netzwerk zugeordnet sind, wobei insbesondere vorgesehen ist, dass das neuronale Netzwerk mit Soll-Trainingsdaten trainiert wird, wobei die Soll-Trainingsdaten Audio-Eingangssignale (a1, a2) und dazu korrespondierende vordefinierte Audioobjekte (16) umfassen, mit den folgenden Trainingsschritten:

   - Vorwärtsspeisen (V15) des neuronalen Netzwerks mit den Soll-Trainingsdaten unter Erhalt eines ermittelten Audioobjekts (17),
   - Bestimmen (V16) eines Fehlervektors (P) zwischen dem ermittelten Audioobjekt (17) und dem vordefinierten Audioobjekt (16) und
   - Ändern von Parametern des neuronalen Netzwerks durch Rückwärtsspeisen (V18) des neuronalen Netzwerks mit dem Fehlervektor (P), falls ein Qualitätsparameter des Fehlervektors (P) einen vordefinierten Wert übersteigt.

10. Verfahren nach einem der Ansprüche 1 bis 9, **dadurch gekennzeichnet, dass** das Verfahren derart ausgestaltet ist, dass es kontinuierlich abläuft und/oder dass die Audio-Eingangssignale (a1, a2) jeweils Teile von insbesondere kontinuierlich eingelesenen Audio-Signalen (b1, b2) mit insbesondere vordefinierten zeitlichen Längen sind.

11. System (10) zur Extraktion eines Audioobjektes (11) aus mindestens zwei Audio-Eingangssignalen (a1, a2) mit einer Steuereinheit (15), die dazu ausgebildet ist, ein Verfahren nach einem der Ansprüche 1 bis 10 durchzuführen.

12. System nach Anspruch 11, **dadurch gekennzeichnet, dass** ein erstes Mikrofon (13) zum Empfangen des ersten Audio-Eingangssignals (a1) und ein zweites Mikrofon (14) zum Empfangen des zweiten Audio-Eingangssignals (a2) jeweils mit dem System (10) derart verbindbar sind, dass die Audio-Eingangssignale (a1, a2) der Mikrofone (13, 14) der Steuereinheit (15) zuführbar sind.

13. System nach einem der Ansprüche 11 oder 12, **dadurch gekennzeichnet, dass** das System (10) als Komponente eines Mischpults (10_a) ausgestaltet ist.

14. Computerprogramm mit Programmcodemitteln, das dazu ausgestaltet ist, die Schritte eines Verfahrens nach einem der Ansprüche 1 bis 10 durchzuführen, wenn das Computerprogramm auf einem Computer oder einer entsprechenden Recheneinheit ausgeführt wird, insbesondere auf einer Steuereinheit (15) eines Systems (10) nach einem der Ansprüche 11 bis 13.

Claims

1. Method for extracting at least one audio object (11) from at least two audio input signals (a1, a2), each of which contains the audio object (11), the method comprising the following steps:

   - synchronizing (V1) the second audio input signal (a2) with the first audio input signal (a1) to obtain a synchronized second audio input signal (a2'),
   - extracting (V2) the audio object (11) by applying at least one trained model to the first audio signal (a1) and to the synchronized second audio input signal (a2'), and
   - outputting (V3) the audio object (11),

   wherein the method step of synchronizing (V1) the second audio input signal (a2) with the first audio input signal (a1) comprises the following method steps:

   - generating (V4) audio signals (m1, m2) by applying a first trained operator to the audio input signals (a1, a2),
   - analytically calculating (V5) a correlation between the audio signals (m1, m2) to obtain a correlation vector (k),
   - optimizing (V6) the correlation vector (k) using a second trained operator to obtain a synchronization vector (s) and
   - determining (V7) the synchronized second audio input signal (a2') using the synchronization vector (s),

   characterized in that the second trained operator has an iterative method with a finite number of iteration steps (I).

2. Method according to claim 1, **characterized in that** the first trained operator comprises a transformation, in particular a trained transformation, of the audio input signals (a1, a2) into a feature space.

3. Method according to either claim 1 or 2, **characterized in that** the second trained operator comprises at least one normalization of the correlation vector (k).

4. Method according to any of claims 1 to 3, **characterized in that,** in each iteration step, a synchronization vector (s) is determined.

5. Method according to claim 4, **characterized in that** the number of iteration steps (I) of the second trained operator can be defined on the user side.

6. Method according to any of claims 4 to 5, **characterized in that,** in each iteration step, a normalization of the synchronization vector (s) and/or a stretched convolution of the synchronized audio input signal (a2') with the synchronization vector (s') takes place.

7. Method according to any of claims 1 to 6, **characterized in that** the second trained operator provides the determination of at least one acoustic model function (M).

8. Method according to any of claims 1 to 7, **characterized in that** the trained model for extracting (V2) the audio object (11) provides at least one transformation of the first audio input signal (a1) and the synchronized second audio input signal (a2'), each into a representation space, in particular a higher-dimensional representation space **and/or in that** the trained model for extracting (V2) the audio object (11) provides the application of at least one learned filter mask to the first audio input signal (a1) and to the synchronized second audio input signal (a2'), wherein, in particular, the trained model for extracting (V2) the audio object (11) provides at least one transformation of the audio object (11) into the time period of the audio input signals (a1, a2).

9. Method according to any of claims 1 to 8, **characterized in that** the method steps of synchronizing (V1) and/or extracting (V2) and/or outputting (V3) the audio object (11) are assigned to a single neural network, wherein, in particular, the neural network is trained with target training data, wherein the target training data comprise audio input

signals (a1, a2) and corresponding predefined audio objects (16), the method comprising the following training steps:

- forward feeding (V15) the target training data to the neural network to obtain an ascertained audio object (17),
- determining (V16) an error vector (P) between the ascertained audio object (17) and the predefined audio object (16), and
- changing parameters of the neural network by backward feeding (V18) the error vector (P) to the neural network if a quality parameter of the error vector (P) exceeds a predefined value.

10. Method according to any of claims 1 to 9, **characterized in that** the method is designed in such a way that it runs continuously **and/or in that** the audio input signals (a1, a2) are each parts of audio signals (b1, b2), which audio signals are in particular continuously read and have in particular predefined time lengths.

11. System (10) for extracting an audio object (11) from at least two audio input signals (a1, a2), which system has a control unit (15) which is configured to carry out a method according to any of claims 1 to 10.

12. System according to claim 11, **characterized in that** a first microphone (13) for receiving the first audio input signal (a1) and a second microphone (14) for receiving the second audio input signal (a2) can each be connected to the system (10) in such a way that the audio input signals (a1, a2) of the microphones (13, 14) can be supplied to the control unit (15).

13. System according to either claim 11 or 12, **characterized in that** the system (10) is configured as a component of a mixing console (10a).

14. Computer program having program code means, which computer program is configured to carry out the steps of a method according to any of claims 1 to 10 when the computer program is run on a computer or a corresponding computing unit, in particular on a control unit (15) of a system (10) according to any of claims 11 to 13.

Revendications

1. Procédé permettant l'extraction d'au moins un objet audio (11) à partir d'au moins deux signaux d'entrée audio (a1, a2) qui contiennent respectivement l'objet audio (11), comportant les étapes suivantes :

- synchronisation (V1) du second signal d'entrée audio (a2) avec le premier signal d'entrée audio (a1) pour obtenir un second signal d'entrée audio synchronisé (a2'),
- extraction (V2) de l'objet audio (11) par l'application d'au moins un modèle entraîné sur le premier signal d'entrée audio (a1) et sur le second signal d'entrée audio synchronisé (a2'), et
- sortie (V3) de l'objet audio (11),

dans lequel l'étape de procédé de synchronisation (V1) du second signal d'entrée audio (a2) avec le premier signal d'entrée audio (a1) comprend les étapes de procédé suivantes :

- génération (V4) de signaux audio (m1, m2) en appliquant un premier opérateur entraîné sur les signaux d'entrée audio (a1, a2),
- calcul analytique (V5) d'une corrélation entre les signaux audio (m1, m2) pour obtenir un vecteur de corrélation (k),
- optimisation (V6) du vecteur de corrélation (k) à l'aide d'un second opérateur entraîné pour obtenir un vecteur de synchronisation (s), et
- détermination (V7) du second signal d'entrée audio synchronisé (a2') à l'aide du vecteur de synchronisation (s),

caractérisé en ce que le second opérateur entraîné présente un procédé itératif comportant un nombre fini d'étapes d'itération (I).

2. Procédé selon la revendication 1, **caractérisé en ce que** le premier opérateur entraîné comprend une transformation, en particulier entraînée, des signaux d'entrée audio (a1, a2) dans un espace de caractéristiques.

3. Procédé selon l'une des revendications 1 ou 2, **caractérisé en ce que** le second opérateur entraîné comprend au moins une normalisation du vecteur de corrélation (k).

4. Procédé selon l'une des revendications 1 à 3, **caractérisé en ce qu'un** vecteur de synchronisation (s) est déterminé à chaque étape d'itération.

5. Procédé selon la revendication 4, **caractérisé en ce que** le nombre d'étapes d'itération (I) du second opérateur entraîné peut être défini côté utilisateur.

6. Procédé selon l'une des revendications 4 à 5, **caractérisé en ce qu'à** chaque étape d'itération, une normalisation du vecteur de synchronisation (s) et/ou une convolution étirée du signal d'entrée audio synchronisé (a2') sont effectuées avec le vecteur de synchronisation (s').

7. Procédé selon l'une des revendications 1 à 6, **caractérisé en ce que** le second opérateur entraîné prévoit la détermination d'au moins une fonction de modèle acoustique (M).

8. Procédé selon l'une des revendications 1 à 7, **caractérisé en ce que** le modèle entraîné de l'extraction (V2) de l'objet audio (11) prévoit au moins une transformation du premier signal d'entrée audio (a1) et du second signal d'entrée audio synchronisé (a2') respectivement dans un espace de représentation, en particulier de dimension supérieure, **et/ou en ce que** le modèle entraîné de l'extraction (V2) de l'objet audio (11) prévoit l'application d'au moins un masque filtrant appris sur le premier signal d'entrée audio (a1) et sur le second signal d'entrée audio synchronisé (a2'), dans lequel le modèle entraîné de l'extraction (V2) de l'objet audio (11) est en particulier prévu pour prévoir au moins une transformation de l'objet audio (11) dans la durée des signaux d'entrée audio (a1, a2).

9. Procédé selon l'une des revendications 1 à 8, **caractérisé en ce que** les étapes de procédé de synchronisation (V1) et/ou d'extraction (V2) et/ou de sortie (V3) de l'objet audio (11) sont associées à un seul réseau neuronal, dans lequel le réseau neuronal est en particulier prévu pour être entraîné avec des données d'entraînement de consigne, dans lequel les données d'entraînement de consigne comprennent des signaux d'entrée audio (a1, a2) et des objets audio prédéfinis (16) qui leur correspondent, comportant les étapes d'entraînement suivantes :

   - alimentation avant (V15) du réseau neuronal avec les données d'entraînement de consigne pour obtenir un objet audio établi (17),
   - détermination (V16) d'un vecteur d'erreur (P) entre l'objet audio établi (17) et l'objet audio prédéfini (16), et
   - modification de paramètres du réseau neuronal en alimentant en rétroaction (V18) le réseau neuronal avec le vecteur d'erreur (P) si un paramètre de qualité du vecteur d'erreur (P) franchit une valeur prédéfinie.

10. Procédé selon l'une des revendications 1 à 9, **caractérisé en ce que** le procédé est configuré de telle sorte qu'il se déroule en continu **et/ou en ce que** les signaux d'entrée audio (a1, a2) sont respectivement des parties de signaux audio (b1, b2) lus, en particulier en continu, avec des longueurs temporelles, en particulier prédéfinies.

11. Système (10) permettant l'extraction d'un objet audio (11) à partir d'au moins deux signaux d'entrée audio (a1, a2), comportant une unité de commande (15) configurée pour mettre en œuvre un procédé selon l'une des revendications 1 à 10.

12. Système selon la revendication 11, **caractérisé en ce qu'**un premier microphone (13) destiné à recevoir le premier signal d'entrée audio (a1) et un second microphone (14) destiné à recevoir le second signal d'entrée audio (a2) peuvent être respectivement connectés au système (10) de telle sorte que les signaux d'entrée audio (a1, a2) des microphones (13, 14) peuvent être amenés à l'unité de commande (15).

13. Système selon l'une des revendications 11 ou 12, **caractérisé en ce que** le système (10) est configuré comme un composant d'un pupitre de mélange (10a).

14. Programme informatique comportant des moyens de code de programme, configuré pour mettre en œuvre les étapes d'un procédé selon l'une des revendications 1 à 10, lorsque le programme informatique est exécuté sur un ordinateur ou une unité de calcul correspondante, en particulier sur une unité de commande (15) d'un système (10) selon l'une des revendications 11 à 13.

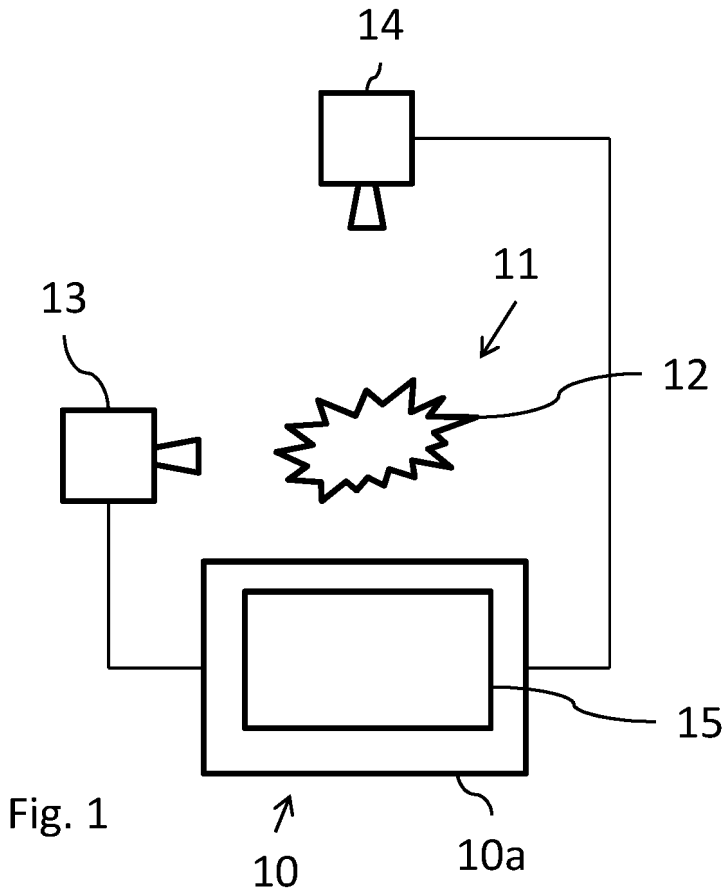


Fig. 1

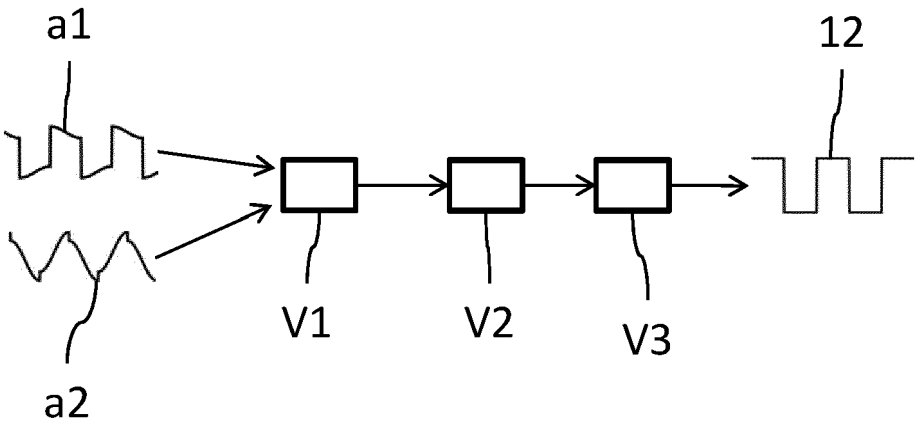


Fig. 2

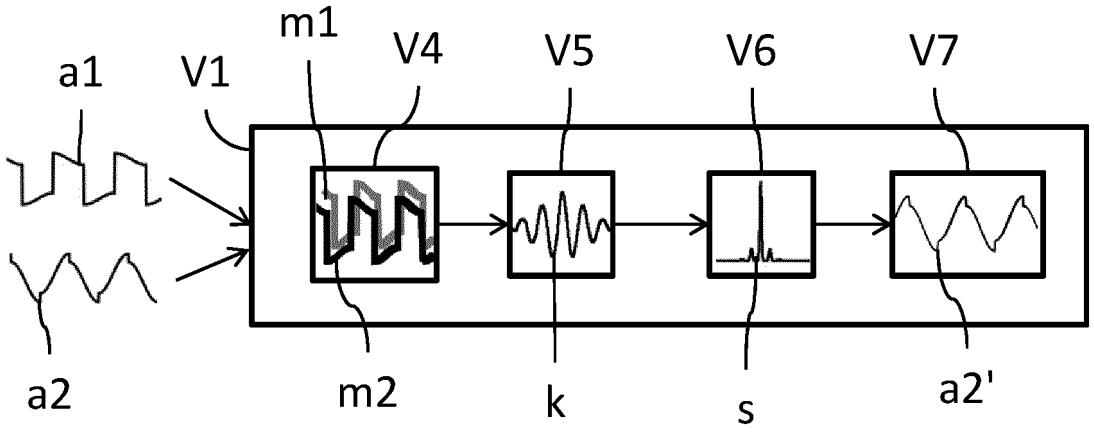


Fig. 3

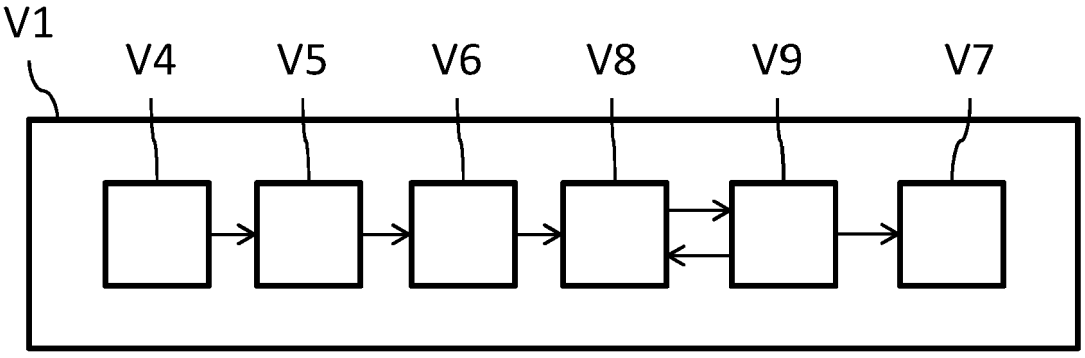


Fig. 4

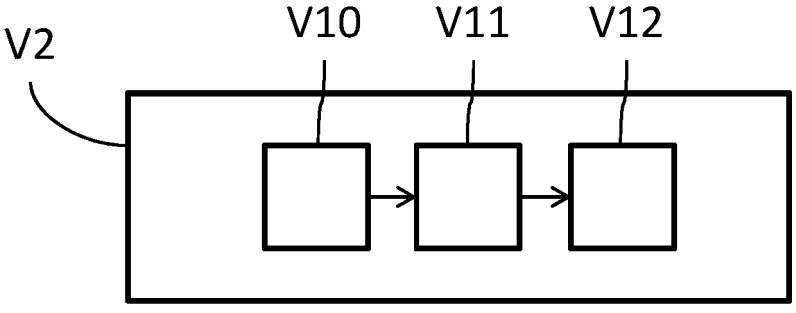


Fig. 5

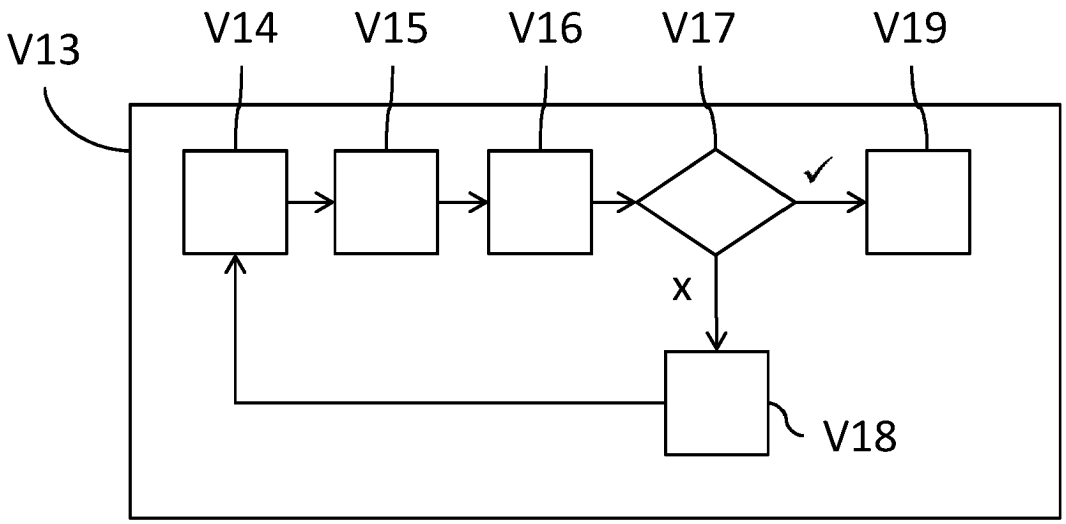


Fig. 6

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.

In der Beschreibung aufgeführte Patentdokumente

- CN 110534127 A **[0006]**

In der Beschreibung aufgeführte Nicht-Patentliteratur

- FaSNet: Low-Latency Adaptive Beamforming for Multi-Microphone Audio Processing. **LUO et al.** AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING WORKSHOP (ASRU). IEEE, 14 December 2019, 260-267 **[0006]**