



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) **DE 698 39 212 T2** 2009.03.19

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 990 370 B1**

(51) Int Cl.<sup>8</sup>: **H04S 3/00** (2006.01)

(21) Deutsches Aktenzeichen: **698 39 212.4**

(86) PCT-Aktenzeichen: **PCT/GB98/01594**

(96) Europäisches Aktenzeichen: **98 925 802.5**

(87) PCT-Veröffentlichungs-Nr.: **WO 1998/058523**

(86) PCT-Anmeldetag: **01.06.1998**

(87) Veröffentlichungstag  
der PCT-Anmeldung: **23.12.1998**

(97) Erstveröffentlichung durch das EPA: **05.04.2000**

(97) Veröffentlichungstag  
der Patenterteilung beim EPA: **05.03.2008**

(47) Veröffentlichungstag im Patentblatt: **19.03.2009**

(30) Unionspriorität:  
**97304218 17.06.1997 EP**

(84) Benannte Vertragsstaaten:  
**DE, DK, FI, FR, GB, NL**

(73) Patentinhaber:  
**British Telecommunications p.l.c., London, GB**

(72) Erfinder:  
**RIMELL, Andrew, Ipswich, Suffolk IP 45BT, GB;  
HOLLIER, Michael Peter, Ipswich, Suffolk IP5 2XX,  
GB**

(74) Vertreter:  
**BEETZ & PARTNER Patentanwälte, 80538  
München**

(54) Bezeichnung: **RAUMKLANGWIEDERGABE**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

## Beschreibung

**[0001]** Diese Erfindung betrifft die Reproduktion bzw. Wiedergabe von räumlichem Audio in Eintauch(immer-sive)-Umgebungen mit nicht idealen akustischen Bedingungen. Eintauch-Umgebungen werden eine wichtige Komponente von zukünftigen Kommunikationssystemen sein. Eine Eintauch-Umgebung ist eine Umgebung, in der dem Benutzer das Empfinden gegeben wird, dass er sich in einer Umgebung befindet, die durch das System dargestellt wird, anstatt sie von außen zu beobachten, wie bei einem herkömmlichen flachen Bildschirm, wie einem Fernseher. Das „Eintauchen“ ermöglicht dem Benutzer, vollständiger in das fragliche Material einbezogen zu sein. Für das visuelle Gefühl kann eine Eintauch-Umgebung erzeugt werden durch derartiges Anordnen, dass das gesamte Sichtfeld des Benutzers mit einer visuellen Darstellung ausgefüllt ist, wodurch ein Eindruck einer Dreidimensionalität entsteht und dem Benutzers ermöglicht wird, eine komplexe Geometrie wahrzunehmen.

**[0002]** Damit der eintauchende Effekt realistisch ist, muss der Benutzer geeignete Eingaben an alle Sinne empfangen, die zu dem Effekt beitragen. Insbesondere ist die Verwendung von kombiniertem Audio und Video ein wichtiger Aspekt der meisten Eintauch-Umgebungen: siehe zum Beispiel:

ANDERSON. D. 85 CASEY. M. „Virtual worlds – The sound dimension“, IEEE Spectrum 1997, Vol. 34, Nr. 3, S. 46–50;

BRAHMAN. R. & COMERFORD. R. „Sharing virtual worlds“, IEEE Spectrum 1997, Vol. 34, Nr. 3, S. 18–20;

WATERS. R. & BARRUS. J. „The rise of shared virtual environments“, IEEE Spectrum 1997, Vol. 34, Nr. 3, S. 20–25.

**[0003]** Räumliches Audio, die Verwendung von zwei oder mehr Lautsprechern, um einen Audioeffekt zu erzeugen, der von dem Zuhörer wahrgenommen wird, als würde er von einer Quelle entstammen, die einen Abstand zu den Lautsprechern hat, ist weithin bekannt. In ihrer einfachsten Form wurden stereofonische Effekte in Audiosystemen für mehrere Jahrzehnte verwendet. In dieser Spezifikation wird der Begriff „virtuelle“ Schallquelle verwendet in der Bedeutung der offensichtlichen Quelle eines Schalls, wie von einem Zuhörer wahrgenommen, im Gegensatz zu den tatsächlichen Schallquellen, welche die Lautsprecher sind.

**[0004]** Eintauch-Umgebungen werden erforscht zur Verwendung in Telepräsenz, Telekonferenzen, „Durchflug“ durch Pläne von Architekten, Erziehung und Medizin. Das weite Sichtfeld kombiniert mit räumlichem Audio erzeugt ein Gefühl eines „Dabeiseins“, was dem Kommunikationsprozess hilft, und die zusätzliche Empfindung von Größe und Tiefe kann einen kraftvollen gemeinschaftlichen Designraum liefern.

**[0005]** Mehrere Beispiele einer Eintauch-Umgebung werden beschrieben von D. M. Traill, J. M. Bowskill und P. J. Lawrence in „Interactive Collaborative Media Environments“ (British Telecommunications Technology Journal, Vol. 15, Nr. 4, Oktober 1997), Seiten 130 bis 139. Ein Beispiel einer eintauchenden Umgebung ist BT/ARC VisionDome (beschrieben auf den Seiten 135 bis 136 und [Fig. 7](#) dieses Artikels), in dem das visuelle Bild auf einem großen konkaven Bildschirm mit den Benutzern im Inneren präsentiert wird (siehe [Fig. 1](#) und [Fig. 2](#)). Ein räumliches Mehrfach-Kanal-Audiosystem mit acht Lautsprechern wird verwendet, um ein Audio-Eintauchen zu liefern. Eine weitere Beschreibung ist zu finden bei:

[http://www.labs.bt.com/people/walkergr/IBTE\\_VisionDome/index.htm](http://www.labs.bt.com/people/walkergr/IBTE_VisionDome/index.htm).

**[0006]** Ein zweites Beispiel ist der „SmartSpace“-Stuhl, der auf den Seiten 134 und 135 (und [Fig. 6](#)) desselben Artikels beschrieben wird, der einen Breitwandvideobildschirm, ein Computerterminal und räumliches Audio kombiniert, die alle angeordnet sind, sich mit der Rotation eines Drehstuhls zu bewegen – ein momentan in Entwicklung stehendes System von British Telecommunications plc. Eine Rotation des Stuhls veranlasst, dass sich die Orientierung des Benutzers in der Umgebung ändert, wobei die visuellen und Audio-Eingaben demgemäß modifiziert werden. Der „SmartSpace“-Stuhl verwendet eine transaurale Verarbeitung, wie beschrieben wird von COOPER. D. & BAUCK. J. „Prospects for transaural recording“, Journal of the Audio Engineering Society 1989, Vol. 37, Nr. 1/2, S. 3–19, um eine „Klangwolke“ um den Benutzer herum zu liefern, was ihm das Gefühl eines vollständigen klanglichen Eintauchens gibt, während der umlaufende Bildschirm ein visuelles Eintauchen liefert.

**[0007]** Wenn die Eintauch-Umgebung interaktiv ist, werden Bilder und räumlicher Schall in Echtzeit erzeugt (typischerweise als eine Computeranimation), während ein nicht interaktives Material mit einer ambisonischen B-Format-Tonspur geliefert wird, wobei deren Charakteristiken später in dieser Spezifikation beschrieben werden. Eine Ambisonic-Codierung ist eine gängige Wahl für Eintauch-Audio-Umgebungen, da es möglich ist, jede Anzahl von Kanälen zu decodieren unter Verwendung von nur drei oder vier Übertragungskanälen. Jedoch hat eine Ambisonic-Technologie ihre Grenzen, wenn sie in Telepräsenz-Umgebungen verwendet wird,

wie diskutiert wird.

**[0008]** Mehrere Fragen hinsichtlich einer Schalllokalisierung in Eintauch-Umgebungen werden nun betrachtet. Die [Fig. 1](#) und [Fig. 2](#) zeigen eine Draufsicht und einen Seitenquerschnitt des VisionDomes mit acht Lautsprechern (1, 2, 3, 4, 5, 6, 7, 8), dem umlaufenden Bildschirm und typischen Benutzerpositionen. Mehrfach-Kanal-Ambisonic-Audiospuren werden typischerweise in rechteckigen Zuhörräumen reproduziert. Wenn in einem hemisphärischen wiedergegeben, wird eine Spatialisierung bzw. Räumlichkeit von der Geometrie der Zuhörumgebung beeinflusst. Reflexionen in der Hemisphäre können die Schallfeld-Neukombination zerstören: obwohl dies manchmal minimiert werden kann durch Behandeln der Wandoberflächen mit einem geeigneten absorptiven Material, muss dies nicht immer praktisch sein. Die Verwendung einer Hartplastikkuppel als Zuhörumgebung erzeugt viele akustische Probleme, die hauptsächlich von mehrfachen Reflexionen verursacht werden. Die akustischen Eigenschaften der Kuppel, wenn unbehandelt, verursachen, dass Schall erscheint, als stamme er von mehreren Quellen, und somit wird der beabsichtigte Klangräumlichkeitseffekt zerstört. Eine Lösung ist, die innere Oberfläche der Kuppel mit einem absorbierenden Material abzudecken, das Reflexionen reduziert. Das Material des Videobildschirms selbst ist Schall-absorbierend, somit hilft es bei der Reduzierung von Schallreflexionen, verursacht aber eine beträchtliche Hochfrequenzdämpfung für Töne, die von Lautsprechern kommen, die sich hinter dem Bildschirm befinden. Diese Hochfrequenzdämpfung wird über wunden durch Anwenden einer Entzerrung auf die Signale, die den Lautsprechern 1, 2, 3, 7, 8 zugeführt werden, die sich hinter dem Bildschirm befinden.

**[0009]** Andere Zuhörumgebungen als eine Plastikkuppel haben ihre eigenen akustischen Eigenschaften und in den meisten Fällen sind Reflexionen eine Fehlerursache. Wie bei einer Kuppel reduziert die Anwendung von Akustikplatten die Menge an Reflexionen, wodurch die Fähigkeit des Benutzers erhöht wird, Audiosignale genau zu lokalisieren.

**[0010]** Die meisten Projektionsbildschirme und Videomonitor haben einen flachen (oder fast flachen) Bildschirm. Wenn eine vorher aufgenommene B-Format-Tonspur erstellt wird, mit einem sich bewegenden Videobild zu übereinstimmen, wird sie typischerweise in Studios mit solchen flachen Videobildschirmen erstellt. Um die richtige räumliche Wahrnehmung (wahrgenommenes Schallfeld) zu liefern, bringt die verwendete B-Format-Codierung das Audio mit dem flachen Videobildschirm in Übereinstimmung. Wenn jedoch große Umgebungen mit mehreren Benutzern verwendet werden, wie der VisionDome, wird das Video auf einem konkaven Bildschirm wiedergegeben, wobei das Videobild geeignet modifiziert wird, um für einen Beobachter korrekt zu erscheinen. Jedoch ist die Geometrie des Audioeffekts nicht langer konsistent mit dem Video und eine nicht-lineare Abbildung ist erforderlich, um die Wahrnehmungssynchronisierung wieder herzustellen. In dem Fall eines interaktiven Materials platziert der B-Format-Codierer die virtuelle Quelle auf den Umfang eines Einheitskreises, wodurch die Krümmung des Bildschirms abgebildet wird.

**[0011]** In Umgebungen, in denen sich eine Gruppe von Zuhörern in einem kleinen Bereich befindet, gelingt es einem Ambisonic-Wiedergabesystem wahrscheinlich nicht, für die meisten die gewünschte Hörräumlichkeit zu erzeugen. Ein Grund ist, dass die verschiedenen Schallfelder, die von den Lautsprechern erzeugt werden, sich nur an einer Position richtig kombinieren, um den gewünschten Effekt einer „virtuellen“ Schallquelle zu erzeugen, als der „Sweetspot“ bzw. ideale Punkt bekannt. Nur ein Zuhörer (höchstens) kann sich in dem genauen Sweetspot befinden. Dies ist aufgrund dessen, da der wahre Sweetspot, wo phasengleiche und gegenphasige Signale richtig rekonstruiert werden, um das gewünschte Signal zu liefern, ein kleiner Bereich ist, und Teilnehmer außerhalb des Sweetspots empfangen eine inkorrekte Kombination von phasengleichen und gegenphasigen Signalen. Tatsächlich ist bei einem hemisphärischen Bildschirm der Videoprojektor normalerweise an dem geometrischen Zentrum der Hemisphäre und die Ambisonics sind im Allgemeinen derart angeordnet, dass der „Sweetspot“ auch an dem geometrischen Zentrum des Lautsprecher-Arrays ist, das konzentrisch mit dem Bildschirm angeordnet ist. Somit kann keiner an dem tatsächlichen Sweetspot sein, da diese Position von dem Projektor besetzt ist.

**[0012]** Der Effekt des Bewegens des Sweetspots, um mit der Position von einem der Zuhörer zusammenzufallen, wurde untersucht von BURRASTON, HOLLIER & HAWKSFORD („Limitations of dynamically controlling the listening position in a 3-D ambisonic environment“, Preprint from 102<sup>nd</sup> AES Convention, März 1997, Audio Engineering Society (Preprint Nr. 4460)). Dies ermöglicht einem Zuhörer, der sich nicht in dem ursprünglichen Sweetspot befindet, die richtige Kombination von ambisonischen decodierten Signalen zu empfangen. Jedoch ist dieses System nur für einzelne Benutzer ausgebildet, da der Sweetspot nur jeweils an eine Position verschoben werden kann. Die Veröffentlichung diskutiert die Effekte eines Zuhörers, der außerhalb des Sweetspots positioniert ist (wie es bei einer Gruppe von Benutzern an einem virtuellen Treffpunkt passiert), und folgert basierend auf zahlreichen formellen Hörtests, dass Zuhörer den Schall nur korrekt lokalisieren können,

wenn sie sich in dem Sweetspot befinden.

**[0013]** Wenn sich eine Schallquelle bewegt und der Zuhörer befindet sich nicht an einer Sweetspot-Position, werden interessante Effekte beobachtet. Es wird ein Beispiel betrachtet, in dem sich der Schall von vorne rechts zu vorne links bewegt und der Benutzer befindet sich nicht in der Mitte und nah vorne. Der Schall scheint anfangs von dem rechten Lautsprecher zu kommen, bleibt dort für eine Weile und bewegt sich dann schnell über die Mitte zu dem linken Lautsprecher – der Schall tendiert dazu, „um den Lautsprecher zu hängen“, wodurch ein akustisch hohler Zentrumsbereich oder „Loch“ entsteht. Für Zuhörer, die sich nicht an dem Sweet-spot befinden, erscheint jede virtuelle Schallquelle im Allgemeinen als zu nahe an einem der Lautsprecher. Wenn sie sich langsam durch den Raum bewegt (wie von einem Zuhörer an dem Sweetspot wahrgenommen), nehmen Benutzer, die sich nicht an dem Sweetspot befinden, die virtuelle Quelle wahr als nahe an einer Lautsprecherposition und dann plötzlich zu einem anderen Lautsprecher springend.

**[0014]** Das einfachste Verfahren einer geometrischen Koordinaten-Korrektur umfasst ein Verzerren der geometrischen Positionen der Lautsprecher, wenn Lautsprecherpositionen in den ambisonischen Decoder programmiert werden. Der Decoder ist programmiert für Lautsprecherpositionen, die näher an der Mitte sind als ihre tatsächlichen Positionen: dies führt zu einem Effekt, in dem sich der Schall an den Rändern des Bildschirms schnell bewegt und langsam in dem Bereich in der Mitte des Bildschirms – was zu einer wahrgenommenen linearen Bewegung des Schalls in Bezug zu einem Bild auf dem Bildschirm führt. Dieses Prinzip kann nur auf ambisonische Decoder angewendet werden, die das B-Format-Signal an wählbare Lautsprecherpositionen decodieren können, d. h. es kann nicht mit Decodern verwendet werden, die für feste Lautsprecherpositionen vorgesehen sind (wie die acht Ecken eines Würfels oder vier Ecken eines Quadrats).

**[0015]** Eine nicht-lineare Schwenkungs(panning)-Strategie wurde entwickelt, die als ihre Eingabe die mono-phone Schallquelle, die gewünschte Schallposition (x, y, z) und die Positionen der N Lautsprecher in dem Wiedergabesystem (x, y, z) nimmt. Dieses System kann jede Anzahl von getrennten Eingangsquellen haben, die individuell zu getrennten Punkten im Raum lokalisiert werden können. Eine virtuelle Schallquelle wird von einer Position zu einer anderen geschwenkt mit einer nicht-linearen Schwenkungscharakteristik. Das nicht-lineare Schwenken korrigiert die oben beschriebenen Effekte, in denen ein Audio-„Loch“ wahrgenommen wird. Das Wahrnehmungserlebnis wird korrigiert, um eine lineare Audio-Trajektorie von ursprünglicher Position zu Endposition zu liefern. Das nicht-lineare Schwenkungs- bzw. Panning-Schema basiert auf einem Intensitäts-Schwenken und nicht auf einer Wellenfront-Wiederherstellung, wie in einem Ambisonic-System. Da die Verzerrung auf einem Intensitäts-Schwenken basiert, gibt es kein gegenphasiges Signal von den anderen Lautsprechern, und somit erfahren alle Zuhörer mit einem Mehrfach-Benutzer-System ein korrekt räumliches Audio. Der nicht-lineare Verzerrungs-Algorithmus ist ein vollständiges System (d. h. es nimmt die Koordinaten eines Signals und positioniert sie in einem dreidimensionalen Raum), somit kann er nur für Echtzeit-Material verwendet werden und nicht zur Verzerrungs-Ambisonic-Aufzeichnungen.

**[0016]** Gemäß der vorliegenden Erfindung ist vorgesehen ein Verfahren zur Erzeugung eines Schallfelds aus einem Array von Lautsprechern, wobei das Array einen Hörraum definiert, in dem die Ausgaben der Lautsprecher kombiniert werden, um eine räumliche Wahrnehmung einer virtuellen Schallquelle zu liefern, wobei das Verfahren die Erzeugung einer jeweiligen Ausgabekomponente  $P_n$  für jeden Lautsprecher in dem Array zum Steuern der Ausgabe des jeweiligen Lautsprechers aufweist, wobei die Ausgabe aus Daten abgeleitet wird, die von einem Eingangssignal getragen werden, wobei die Daten ein Summenreferenzsignal W und Richtungsschallkomponenten X, Y, (Z) aufweisen, welche die Schallkomponente in unterschiedlichen Richtungen, wie durch die virtuelle Schallquelle erzeugt, repräsentieren, wobei das Verfahren die Schritte aufweist des Erkennens für jeden Lautsprecher, ob die jeweilige Komponente  $P_n$  sich in Phase oder Gegenphase zu dem Summenreferenzsignal W ändert, des Modifizierens des Signals, wenn es sich in Gegenphase befindet, und des Zuführens der resultierenden modifizierten Komponenten zu den jeweiligen Lautsprechern.

**[0017]** Gemäß einem zweiten Aspekt der Erfindung ist vorgesehen eine Vorrichtung zur Erzeugung eines Schallfelds, die aufweist ein Array von Lautsprechern, die einen Hörraum definieren, wobei die Ausgaben der Lautsprecher kombiniert werden, um eine räumliche Wahrnehmung einer virtuellen Schallquelle zu liefern, Mittel zum Empfang und zur Verarbeitung von Daten, die von einem Eingangssignal getragen werden, wobei die Daten ein Summenreferenzsignal W und Richtungsschallkomponenten X, Y, (Z) aufweisen, welche den Schall in unterschiedlichen Richtungen, wie durch die virtuelle Schallquelle erzeugt, anzeigen, Mittel zur Erzeugung einer jeweiligen Ausgabekomponente  $P_n$  zur Steuerung der Ausgabe jedes Lautsprechers in dem Array aus diesen Daten, Mittel zur Erkennung für jeden Lautsprecher, ob die jeweilige Komponente  $P_n$  sich in Phase oder Gegenphase zu dem Summenreferenzsignal W ändert, Mittel zum Modifizieren des Signals, wenn es sich in Gegenphase befindet, und Mittel zum Zuführen der resultierenden modifizierten Komponenten zu den jewei-

gen Lautsprechern.

**[0018]** Vorzugsweise werden die Richtungsschallkomponenten jeweils multipliziert mit einem Verzerrungsfaktor, der eine Funktion der jeweiligen Richtungsschallkomponente ist, so dass eine sich bewegende virtuelle Schallquelle, die einer gleichmäßigen Trajektorie folgt, wie von einem Zuhörer an einem Punkt in dem Hörfeld wahrgenommen, auch einer gleichmäßigen Trajektorie folgt, wie an einem anderen Punkt in dem Hörfeld wahrgenommen. Dies stellt sicher, dass virtuelle Schallquellen nicht dazu neigen, in bestimmten Bereichen des Hörfelds öfter als andere aufzutreten. Der Verzerrungsfaktor kann eine quadratische oder höhere geradzahlige Potenz oder eine Sinusfunktion der Richtungsschallkomponente sein.

**[0019]** Die ambisonischen B-Format-Codierungs- und -Decodierungs-Gleichungen für zweidimensionale Wiedergabesysteme werden nun kurz diskutiert. Dieser Abschnitt diskutiert nicht die detaillierte Theorie von Ambisonics, sondern zeigt die Ergebnisse von anderen Forschern auf diesem Gebiet. Die ambisonische Theorie liefert eine Lösung für das Problem der Codierung einer Richtungsinformation in ein Audiosignal. Das Signal soll wiedergegeben werden über ein Array aus zumindest vier Lautsprechern (für ein pantophonisch-„horizontale Ebene“-System) oder acht Lautsprechern (für ein periphonisch-„horizontale und vertikale Ebene“-System). Das Signal, als „B-Format“ bezeichnet, besteht (für den ersten Fall) aus drei Komponenten für pantophonische Systeme (W, X, Y) und vier Komponenten für periphonische Systeme (W, X, Y, Z). Für eine detaillierte Analyse von Raumklang und ambisonischer Theorie siehe:

BAMFORD. J. & VANDERKOOY. J. "Ambisonic sound for us" Preprint from 99th AES Convention October 1995 Audio Engineering Society (Preprint Nr. 4138).

BEGAULT. D. "Challenges to the successful implementation of 3-D sound" Journal of the Audio Engineering Society 1991, Vol. 39, Nr. 11, S. 864–870.

BURRSTON et al (Bezugnahme oben).

GERZON. M. "Optimum reproduction matrices for multi-speaker stereo" Journal of the Audio Engineering Society 1992, Vol. 40, Nr. 7/8, S. 571–589.

GERZON. M. "Surround sound psychoacoustics" Wireless World December 1974, Vol. 80, S. 483–485.

MALHAM. D. G. "Computer control of ambisonic soundfields" Preprint from 82nd AES Convention March 1987 Audio Engineering Society (Preprint Nr. 2463).

MALHAM. D. G. & CLARKE. J. "Control software for a programmable soundfield controller" Proceedings of the Institute of Acoustics Autumn Conference on Reproduced Sound 8, Windermere 1992, S. 265–272.

MALHAM. D. G. & MYATT. A. "3-D Sound spatialisation using ambisonic techniques" Computer Music Journal 1995, Vol. 19 Nr. 4, S. 58–70.

POLETTI. M. "The design of encoding functions for stereophonic and polyphonic sound systems" Journal of the Audio Engineering Society 1996, Vol. 44, Nr. 11, S. 948–963.

VANDERKOOY. J. & LIPSHITZ. S. "Anomalies of wavefront reconstruction in stereo and surround-sound reproduction" Preprint from 83rd AES Convention October 1987 Audio Engineering Society (Preprint Nr. 2554).

**[0020]** Die hier beschriebenen ambisonischen Systeme sind alle erster Ordnung, d. h.  $m = 1$ , wobei die Anzahl von Kanälen gegeben wird durch  $2m + 1$  für ein zweidimensionales System (3 Kanäle: w, x, y) und  $(m + 1)^2$  für ein dreidimensionales System (4 Kanäle: w, x, y, z). In dieser Spezifikation werden nur zweidimensionale Systeme betrachtet, jedoch können die hier gezeigten Ideen einfach skaliert werden zur Verwendung mit einem vollständigen dreidimensionalen Wiedergabesystem und der Umfang der Ansprüche umfasst derartige Systeme.

**[0021]** In einem zweidimensionalen System ist der codierte räumliche Schall nur in einer Ebene, der (x, y)-Ebene. Es wird angenommen, dass die Schallquelle in einem Einheitskreis positioniert ist, d. h.  $x^2 + y^2 \leq 1$  (siehe [Fig. 3](#)). Für ein monophonisches Signal, das auf dem Einheitskreis positioniert ist:

$$x = \cos(\varphi)$$

$$y = \sin(\varphi)$$

wobei  $\varphi$  der Winkel zwischen dem Ursprung und der gewünschten Position der Schallquelle ist, wie in der [Fig. 3](#) definiert.

**[0022]** Das B-Format-Signal weist drei Signale W, X, Y auf, die definiert sind als (siehe die Bezugnahme auf Malham und Myatt oben):

$$W = S \cdot \frac{1}{\sqrt{2}}$$

$$X = S \cdot \cos(\varphi)$$

$$Y = S \cdot \sin(\varphi)$$

wobei S das monophonische Signal ist, das räumlich anzuordnen ist.

**[0023]** Wenn die virtuelle Schallquelle auf dem Einheitskreis ist;  $x = \cos(\varphi)$  und  $y = \sin(\varphi)$ , sind somit Gleichungen für W, X, Y hinsichtlich x & y:

$$W = \frac{1}{\sqrt{2}} \cdot S \quad \text{Umgebungssignal}$$

$$X = x \cdot S$$

Vorne-Hinten-Signal

$$Y = y \cdot S$$

Links-Rechts-Signal

**[0024]** Wie ebenfalls von Malham und Myatt beschrieben, arbeitet der Decoder wie folgt. Für ein reguläres Array von N Lautsprechern ist die Decodiergleichung des pantophonischen Systems:

$$P_n = \frac{1}{N} W + 2X \cos(\varphi_n) + 2Y \sin(\varphi_n)$$

wobei  $\varphi_n$  die Richtung des Lautsprechers „n“ (siehe [Fig. 4](#)) ist, und somit sind für ein reguläres Array mit vier Lautsprechern, wie in [Fig. 4](#) gezeigt, die Signale, die den jeweiligen Lautsprechern zugeführt werden:

$$P_1 = \frac{1}{4} W + \frac{2X}{\sqrt{2}} + \frac{2Y}{\sqrt{2}} \quad P_2 = \frac{1}{4} W - \frac{2X}{\sqrt{2}} + \frac{2Y}{\sqrt{2}}$$

$$P_4 = \frac{1}{4} W + \frac{2X}{\sqrt{2}} - \frac{2Y}{\sqrt{2}} \quad P_3 = \frac{1}{4} W - \frac{2X}{\sqrt{2}} - \frac{2Y}{\sqrt{2}}$$

**[0025]** Es ist möglich, unter Verwendung des Verfahrens der Erfindung, ein ambisonisches B-Format-Signal (oder ein verzerrtes B'-Format-Signal, wird noch beschrieben) zu nehmen und die Gegenphase-Komponente zu reduzieren, wodurch ein nicht-lineares Signal des Schwenkungs-Typs erzeugt wird, was einer Gruppe von Benutzern ermöglicht, einen räumlichen Klang zu erleben. Die Wiedergabe ist nicht länger ein ambisonisches System, da eine wahre Wellenfront-Wiederherstellung nicht mehr erreicht wird. Der Decoder-Verzerrungs-Algorithmus nimmt die Ausgaben von dem ambisonischen Decoder und verzerrt sie, bevor sie jedem Wiedergabekanal zugeführt werden, somit gibt es eine Implementierung des Decoder-Verzerrers für jeden der N Ausgabekanäle. Wenn das Signal von einer der B-Format- oder B'-Format-Decoder-Ausgaben eine phasenverschobene Komponente ist, wird ihre Phase umgekehrt hinsichtlich dem W Eingangssignal – somit ist es durch Vergleichen der Decoder-Ausgaben mit W möglich, festzustellen, ob das Signal phasenverschoben ist oder nicht. Wenn eine gegebene Decoder-Ausgabe phasenverschoben ist, dann wird diese Ausgabe um den Dämpfungsfaktor D gedämpft:

$$P'_n = P_n \cdot D$$

wobei  $0 \leq D \leq 1$ , wenn Vorzeichen ( $P_n$ )  $\neq$  Vorzeichen (W), und ansonsten  $D = 1$ .

**[0026]** Der einfache Algorithmus reduziert die Wahrscheinlichkeit, dass eine Schalllokalisierung auf den nächsten Lautsprecher reduziert wird, wenn der Zuhörer weg von dem Sweetspot ist.

**[0027]** Eine B-Format-Verzerrung nimmt eine ambisonische B-Format-Aufzeichnung und korrigiert für die wahrgenommene nicht-lineare Trajektorie. Die Eingabe in das System ist die B-Format-Aufzeichnung und die Ausgabe ist eine verzerrte B-Format-Aufzeichnung (hier als B'-Format-Aufzeichnung bezeichnet). Die B'-For-

mat-Aufzeichnung kann decodiert werden mit jedem B-Format-Decoder, was die Verwendung von existierenden Decodern ermöglicht. Ein ambisonisches System erzeugt einen „Sweetspot“ in dem Wiedergabebereich, wo das Schallfeld korrekt wiederhergestellt ist, und in anderen Bereichen erfahren die Zuhörer keinen richtig lokalisierten Klang. Das Ziel des Verzerrungs-Algorithmus ist, von einem linearen Bereich von  $x$  &  $y$ -Werten zu einem nicht-linearen Bereich zu wechseln. Man betrachte das Beispiel, wenn sich ein Schall von rechts nach links bewegt; der Schall muss sich zuerst schnell bewegen, dann langsam durch die Mitte und schließlich schnell über die linke Seite, um eine korrekte Wahrnehmung zu liefern. Eine Verzerrung betrifft auch die Wahrnehmungssicht von stationären Objekten, da ohne eine Verzerrung Zuhörer, die sich entfernt von dem Sweet-spot befinden, die meisten virtuellen Schallquellen als konzentriert in wenigen Bereichen wahrnehmen, wobei der zentrale Bereich typischerweise weniger besetzt ist und als Audio-„Loch“ wahrgenommen wird. Angesichts der B-Format-Signal-Komponenten  $X$ ,  $Y$  &  $Z$  ist es möglich, Schätzungen der ursprünglichen Werte von  $x$  &  $y$  zu bestimmen, so kann das ursprüngliche Signal  $S$  wiederhergestellt werden, um  $S' = W\sqrt{2}$  zu liefern, woraus die Schätzungen  $x'$  &  $y'$  zu finden sind:

$$x' = \frac{X}{S'} \quad \text{und} \quad y' = \frac{Y}{S'}$$

$$\text{so } x' = \frac{X}{W\sqrt{2}} \quad \text{und} \quad y' = \frac{Y}{W\sqrt{2}}$$

Es sollen  $\hat{x}'$  und  $\hat{y}'$  normalisierte  $x$ - und  $y$ -Werte in dem Bereich  $(\pm 1, \pm 1)$  darstellen. Ein allgemeiner Verzerrungsalgorithmus wird gegeben durch:

$$X' = X \cdot f(\hat{x}') \quad \text{und} \quad Y' = Y \cdot f(\hat{y}')$$

**[0028]** Wenn jedoch  $x$  eine Funktion von  $X$  und  $y$  eine Funktion von  $Y$  ist, dann

$$X' = X \cdot f(X) \quad \text{und} \quad Y' = Y \cdot f(Y)$$

**[0029]** Das resultierende Signal  $X'$ ,  $Y'$  &  $W$  wird als das B'-Format-Signal bezeichnet.

**[0030]** Zwei mögliche Verzerrungsfunktionen werden nun beschrieben.

#### 1) Potenz-Verzerrung

**[0031]** Bei einer Potenz-Verzerrung wird der Wert von  $X$  multipliziert mit  $\hat{x}'$  hoch einer geraden Potenz (effektiv ist  $X$  hoch einer ungeraden Potenz – wodurch das Vorzeichen beibehalten wird),  $Y$  wird auf dieselbe Weise verzerrt.

$$f(x) = (\hat{x}')^{2i} \quad \text{und} \quad f(y) = (\hat{y}')^{2i}$$

$$f(X) = \left( \frac{X}{W\sqrt{2}} \right)^{2i} \quad \text{und} \quad f(Y) = \left( \frac{Y}{W\sqrt{2}} \right)^{2i}$$

$$\text{d.h.} \quad X' = X \cdot \left( \frac{X}{W\sqrt{2}} \right)^{2i} \quad \text{und} \quad Y' = Y \cdot \left( \frac{Y}{W\sqrt{2}} \right)^{2i}$$

**[0032]** In diesen Gleichungen liefert eine Wahl von  $i = 0$  eine nicht-verzerrte Anordnung, während für  $i > 0$  eine nicht-lineare Verzerrung erzeugt wird.

#### 2) Sinus-Verzerrung

**[0033]** Bei einer Sinus-Verzerrung werden unterschiedliche Funktionen,  $f(X)$  &  $f(Y)$ , verwendet für unterschiedliche Teile der  $\hat{x}'$ - und  $\hat{y}'$ -Bereiche. Das Ziel bei einer Sinus-Verzerrung ist, einen konstanten Pegel zu liefern, wenn die virtuelle Schallquelle an den Extremen ihres Bereichs ist, und einen schnellen Übergang zu dem mittigen Bereich. Die Hälfte eines Zyklus' einer erhöhten Sinuswelle wird verwendet, um zwischen den Extremen und dem mittigen Bereich glatt zu interpolieren.

**[0034]** Für  $X$ :



1.  $-1 < \hat{x}' \leq x_1$   $f(X) = \frac{1}{|\hat{x}'|}$
2.  $x_1 < \hat{x}' \leq x_2$   $f(X) = \frac{1}{2 \cdot |\hat{x}'|} \left\{ \sin \left( \frac{(\hat{x}' + |x_1|) \cdot \pi}{|x_2 - x_1|} + \frac{\pi}{2} \right) + 1 \right\}$
3.  $x_2 < \hat{x}' \leq x_3$   $f(X) = 0$
4.  $x_3 < \hat{x}' \leq x_4$   $f(X) = \frac{1}{2 \cdot |\hat{x}'|} \left\{ \sin \left( \frac{(\hat{x}' - |x_3|) \cdot \pi}{|x_4 - x_3|} + \frac{\pi}{2} \right) - 1 \right\}$
5.  $x_4 < \hat{x}' \leq +1$   $f(X) = \frac{1}{|\hat{x}'|}$

Für Y:

1.  $-1 < \hat{y}' \leq y_1$   $f(Y) = \frac{1}{|\hat{y}'|}$
2.  $y_1 < \hat{y}' \leq y_2$   $f(Y) = \frac{1}{2 \cdot |\hat{y}'|} \left\{ \sin \left( \frac{(\hat{y}' + |y_1|) \cdot \pi}{|y_2 - y_1|} + \frac{\pi}{2} \right) + 1 \right\}$
3.  $y_2 < \hat{y}' \leq y_3$   $f(Y) = 0$
4.  $y_3 < \hat{y}' \leq y_4$   $f(Y) = \frac{1}{2 \cdot |\hat{y}'|} \left\{ \sin \left( \frac{(\hat{y}' - |y_3|) \cdot \pi}{|y_4 - y_3|} + \frac{\pi}{2} \right) - 1 \right\}$
5.  $y_4 < \hat{y}' \leq +1$   $f(Y) = \frac{1}{|\hat{y}'|}$

Typische Werte für die Konstanten  $x_{1...4}$  und  $y_{1...4}$  sind:

$$\begin{aligned} x_1 &= y_1 = -0.75; \\ x_2 &= y_2 = -0.25; \\ x_3 &= y_3 = 0.25; \\ x_4 &= y_4 = 0.75; \end{aligned}$$

**[0035]** Die Verwendung eines B-Format-Signals als die Eingabe in den Verzerrungs-Algorithmus hat viele Vorteile gegenüber anderen Techniken. In einer virtuellen Konferenzumgebung kann die Stimme eines Benutzers mit einem B-Format-Signal codiert werden, das dann an alle anderen Benutzer in dem System übertragen wird (die sich überall auf der Welt aufhalten können). Die physikalische Umgebung, in der sich die anderen Benutzer befinden, kann beträchtlich variieren, einer kann ein auf binauralen Kopfhörern basierendes System verwenden (siehe MOLLER, H. "Fundamentals of binaural technology" Applied Acoustics 1992, Vol. 36, S. 171–218). Eine andere Umgebung kann in einem VisionDome unter Verwendung von verzerrten Ambisonics sein. Noch andere können wahre ambisonische Systeme für einzelne Benutzer verwenden oder transaurale Wiedergabesysteme mit zwei Lautsprechern, wie von Cooper und Bauck beschrieben wird (oben angeführt). Dieses Konzept wird in der [Fig. 5](#) gezeigt.

**[0036]** Zwei Implementierungen der Erfindung (eine digital, die andere analog) unter Verwendung einer proprietären Ausrüstung wird im Folgenden beschrieben. In einer virtuellen Konferenzumgebung muss Audio in Echtzeit verarbeitet werden. Es wird hier angenommen, dass es erforderlich ist, dass die gesamte Decodierung in Echtzeit ausgeführt wird unter Verwendung entweder einer analogen oder DSP-basierten Hardware.



**[0037]** Praktische virtuelle Konferenzorte bzw. Treffpunkte können durch einige Meter oder durch viele Tausende von Kilometern getrennt sein. Die Audioverbindungen zwischen jedem Teilnehmer sind typischerweise über digitale Breitbandnetzwerke, wie ISDN, LAN oder WAN. Es ist somit vorteilhaft, die Codierung und Decodierung in der digitalen Domain auszuführen, um unnötige D/A- und A/D-Umwandlungsstufen zu vermeiden. Die Codierung wird ausgeführt durch Verwendung von herkömmlichen B-Format-Codierern und die Decodierung durch einen modifizierten (Verzerrungs)-Decoder. Die Ausnahme dazu ist die Verwendung eines nicht-linearen Schwenkens, das entweder ein monophonisches Signal mit seinen Koordinaten oder ein N-Kanal-Signal senden muss – was ein nicht-lineares Schwenken weniger geeignet macht zur Verwendung in einem System, das entfernte virtuelle Konferenzorte einsetzt.

**[0038]** Der „Lake HURON DSP“-Motor ist ein proprietäres Verfahren zur Erzeugung und Decodierung von ambisonischen B-Format-Signalen, es kann sowohl 2-D- als auch 3-D-Audio mit jeder Anzahl von beliebig angeordneten Lautsprechern decodieren. Eine Beschreibung ist zu finden bei <http://www.lakedsp.com//index.htm>. Der Huron ist ausgestattet mit den erforderlichen Werkzeugen, um kundenspezifische DSP-Programme zu erzeugen und da die mathematischen Grundlagen der Verzerrungs-Algorithmen, die hier gezeigt werden, relativ einfach sind, können sie in einer Implementierung eines ambisonischen Decoders aufgenommen werden. Der Hauptvorteil dieses Verfahrens ist, dass die Hardware bereits entwickelt ist und das System eine große Anzahl von E/A-Kanälen handhaben kann.

**[0039]** Ein zweites Verfahren einer digitalen Implementierung kann ein Programmieren eines DSP-Chips auf einem der vielen DSP-Entwicklungssysteme umfassen, die von den führenden DSP-Chip-Herstellern verfügbar sind. Ein derartiges System erfordert 2 oder 3 Eingabekanäle und eine große Anzahl von Ausgabekanälen (normalerweise vier oder acht). Eine solche Implementierung würde einen hoch spezialisierten Decoder erzeugen, der einfach in großer Zahl hergestellt werden könnte.

**[0040]** Da die Technologie von PCs und Soundkarten zunimmt, wird ein ambisonisches Decodieren und Verzerrern in Echtzeit eine praktische Realität – was die Anforderung für eine komplexe DSP-System-Gestaltung reduziert.

**[0041]** Die B-Format-Verzerrung und die Decoder-Verzerrung können alternativ in der analogen Domain ausgeführt werden unter Verwendung von analogen Multiplizierern. Ein herkömmlicher ambisonischer Decoder kann verwendet werden, um die B'-Format-Decodierung durchzuführen, wobei die Ausgaben des Decoders der Decoder-Verzerrer-Hardware zugeführt werden, ein derartiges System wird in der [Fig. 6](#) gezeigt. Blockdiagramme des B-Format-Verzerrers und des Decoder-Verzerrers werden jeweils in den [Fig. 7](#) und [Fig. 8](#) gezeigt. Die Blockdiagramme entsprechen den Funktionsblöcken, die von analogen Multiplizieren verfügbar sind, der allgemeinen Art, die bei <http://www.analog.com/products/index/12.html> beschrieben werden.

**[0042]** Eine Anzahl von Simulationen unter Verwendung der oben beschriebenen Verfahren wird nun beschrieben. Statt in Echtzeit zu arbeiten, wie es für ein praktisches Ausführungsbeispiel erforderlich wäre, wurde die Verarbeitung, die zur Erzeugung dieser Beispiele verwendet wurde, offline berechnet unter Verwendung eines PCs mit einer geeigneten Audioschnittstelle. Es wird zuerst ein Beispiel betrachtet, in dem eine einzelne Schallquelle von  $(-1, -1)$  zu  $(1, 1)$  bewegt werden soll, unter der Annahme von normalisierten Koordinaten, wobei  $x$  und  $y$  nur jeweils Werte zwischen  $-1$  und  $+1$  annehmen können. Am Anfang der Audiospur befindet sich der virtuelle Schall an der Position  $(-1, -1)$  und am Ende der Spur befindet sich die virtuelle Schallquelle an der Position  $(1, 1)$ . Der Schall ist codiert, sich linear von seiner Startposition zu seiner Endposition zu bewegen. Zur Einfachheit der Darstellung wurde das monophonische Quellensignal, das räumlich angeordnet werden soll, auf eine positive DC-Spannung gesetzt. Unter Verwendung der B-Format-Codier-Technik, die oben beschrieben wird, wurde ein 3-Kanal-Signal konstruiert, das dann mit den Verzerrungsalgorithmen decodiert wurde, ebenfalls oben beschrieben.

**[0043]** [Fig. 9](#) zeigt die Ausgabe jedes der vier Lautsprecher-Zuführen von einem Vier-Kanal-Decoder unter Verwendung einer herkömmlichen ambisonischen B-Format-Codierung, wobei die Lautsprecher-Geometrie in der [Fig. 4](#) gezeigt wird. Es ist zu sehen, dass die virtuelle Quelle sich anfangs in der Nähe des Lautsprechers **3** befindet, der anfangs eine Ausgabe mit voller Größe hat, der Lautsprecher **1** anfangs eine Gegenphase-Ausgabe hat und die Lautsprecher **2** & **4** den Wert von  $W$  haben. Wenn sich die virtuelle Quelle durch den mittigen Bereich bewegt, sind die Pegel der Lautsprecher **1**, **2**, **3** & **4** gleich. Am Ende der beispielhaften Trajektorie hat der Lautsprecher **1** einen hohen Ausgabepiegel, der Lautsprecher **3** ist in der Gegenphase und **2** & **4** bleiben auf dem konstanten  $W$ -Pegel.

**[0044]** [Fig. 10](#) zeigt den Effekt einer Einführung einer B-Format-Verzerrung (ein B'-Format-Signal). Die Lautsprecher haben an den Start- und Endpunkten der Trajektorie ähnliche Pegel zu einer herkömmlichen B-Format-Verzerrung, jedoch befindet sich der Pfad nun hauptsächlich in dem mittleren Bereich, wodurch die Wahrnehmung eliminiert wird, dass der Klang um einzelne Lautsprecher „herum hängt“ („hanging around“) oder sich auf diese „reduziert“ („collapsing to“).

**[0045]** Die Lautsprecher-Zuführen, die in den [Fig. 9](#) und [Fig. 10](#) gezeigt werden, sind für ein ambisonisches Signal – wenn das korrekte Signal an dem Sweetspot durch die Vektorsummierung der phasengleichen und gegenphasigen Signale erlangt wird. Der Decoder-Verzerrungs-Algorithmus dämpft die Gegenphasen-Komponenten, und präsentiert ein kohärenteres Signal für Zuhörer, die sich nicht an dem Sweetspot befinden. Die [Fig. 11](#) zeigt die grundlegende ambisonische B-Format-Decodierung (wie in [Fig. 9](#) zu sehen ist), zusätzlich mit der Anwendung einer Decoder-Verzerrung. Das Entfernen der Gegenphasen-Komponente ist deutlich zu sehen in diesem Beispiel, wenn  $D = 0$ .

**[0046]** Die [Fig. 12](#) zeigt eine B'-Format-Decodierung (wie in [Fig. 10](#) zu sehen ist) mit einer Decoder-Verzerrung, und der Effekt der Gegenphasen-Dämpfung ist zu sehen.

**[0047]** Das obige Beispiel betrachtete eine Trajektorie von  $(-1, -1)$  zu  $(1, 1)$ , d. h. hinten-links nach vorne-rechts: das folgende Beispiel betrachtet eine Trajektorie von  $(1, 1)$  zu  $(-1, -1)$ , d. h. vorne-rechts zu hinten-links. Die [Fig. 13](#), [Fig. 14](#), [Fig. 15](#) und [Fig. 16](#) zeigen jeweils die Effekte des B-Format-Decoders, des B'-Format-Decoders, des B-Format-Decoders mit Decoder-Verzerrung und des B'-Format-Decoders mit Decoder-Verzerrung. In diesem Beispiel ist das Gegenphase-Signal bedeutender aufgrund der gewählten Trajektorie der virtuellen Quelle. Wie bei dem vorhergehenden Beispiel wird der Decoder-Verzerrungsfaktor  $D$  auf Null gesetzt, wodurch alle Gegenphasen-Komponenten entfernt werden.

**[0048]** Zur Einfachheit der graphischen Darstellung verwenden die hier beschriebenen zwei Beispiele eine positive DC-Spannung als die virtuelle Quelle. Jedoch werden in der Praxis Sinuswellen und komplexe Wellenformen (tatsächliche Audiosignale) verwendet. Die Decoder-Algorithmen wurden mit komplexen Wellenformen getestet, um ihren korrekten Betrieb sicherzustellen.

**[0049]** Der letzte Schiedsrichter über die Leistung eines räumlichen Audios ist der Zuhörer. Ein Audio-Klang-Effekt wurde in B-Format-Signale codiert mit einer vorne-rechts zu vorne-links Trajektorie und dann decodiert mit denselben vier Decodier-Algorithmen, die oben beschrieben werden. Formlose Hörtests wurden in dem VisionDome durchgeführt und die folgenden Beobachtungen wurden von den Zuhörern an den folgenden Hörpositionen gemacht:

#### 1. An dem Sweetspot

- B-Format

Die Lautsprechersignale kombinierten korrekt, um die Wahrnehmung einer sich bewegenden Schallquelle zu geben. Jedoch schien, aufgrund der Geometrie und den akustischen Eigenschaften der Hörumgebung, sich der Schall nicht mit einer linearen Trajektorie durch den Hörraum zu bewegen.

- B'-Format

Wie bei dem Beispiel des B-Formats wurden die einzelnen Schallfelder korrekt wiederhergestellt, um die Wahrnehmung einer sich bewegenden Schallquelle zu geben. Die virtuelle Schallquelle hatte eine wahrgenommene lineare Trajektorie aufgrund der Verwendung einer nicht-linearen Verzerrung.

- B-Format mit Decoder-Verzerrung

Der Schall schien sich durch den Hörbereich zu bewegen mit einer nicht-linearen Trajektorie. Die Wahrnehmung war ähnlich zu dem Beispiel des B-Formats.

- B'-Format mit Decoder-Verzerrung

Der Schall schien sich durch den Hörbereich zu bewegen mit einer linearen Trajektorie. Die Wahrnehmung war ähnlich zu dem Beispiel des B'-Formats.

#### 2. In der Nähe der vorne-links- oder vorne-rechts-Lautsprecher (Positionen 1 & 4 in der [Fig. 4](#))

- B-Format

Der Ort der virtuellen Schallquelle „reduziert sich“ auf den nächsten Lautsprecher – der Beitrag dieses Lautsprechers dominiert die Hörumgebung und wenig oder kein Wahrnehmen einer Trajektorie wird erreicht.

- B'-Format

Der Ort der virtuellen Schallquelle „reduziert sich“ auf den nächsten Lautsprecher – der Beitrag dieses Laut-

sprechers dominiert die Hörumgebung, aber es gibt eine geringe Wahrnehmung einer Trajektorie, da das gesamte Schallfeld keinen Beitrag von den hinteren Gegenphase-Lautsprecher-Zuführen hat.

- B-Format mit Decoder-Verzerrung

Eine verbesserte Wahrnehmung einer Bewegung, jedoch ist die wahrgenommene Trajektorie nicht-linear.

- B'-Format mit Decoder-Verzerrung

Eine deutliche Wahrnehmung, dass sich der Schall von einer Position an eine andere bewegt mit einem ungefähr linear wahrgenommenen Trajektoriepfad.

### 3. In der Mitte zwischen den vorne-links- & hinten-links-Lautsprechern (4 & 3) oder in der Mitte zwischen den vorne-rechts- & hinten-rechts-Lautsprechern (1 & 2)

- B-Format

Zwei getrennte Trajektorien werden wahrgenommen: Das phasengleiche Signal (von den Lautsprechern 4 & 1) bewegt sich von rechts nach links und das gegenphasige Signal bewegt sich von links nach rechts. Die zwei getrennten Trajektorien verursachen Verwirrung und tragen mehr zur Ablenkung bei als gar keine Trajektorie.

- B'-Format

Die Wahrnehmung dieses Signals ist ähnlich zu der des B-Format-Signals, aber geringer ausgeprägt – es gab eine geringere Wahrnehmung von zwei getrennten Trajektorien virtueller Quellen.

- B-Format mit Decoder-Verzerrung

Nur eine Trajektorie wurde beobachtet, aber die Trajektorie war deutlich nicht-linear.

- B'-Format mit Decoder-Verzerrung

Hier wurde eine Trajektorie beobachtet, die mehr linear war in ihrer wahrgenommenen Trajektorie als das B'-Format-Signal, ein größerer Grad einer nicht-linearen Verzerrung kann die Lokalisierung noch deutlicher machen.

### 4. Zwischen den hinten-links- & hinten-rechts-Lautsprechern (3 & 2)

- B-Format

Da die zwei dominanten Lautsprecherquellen die hinteren Lautsprecher (2 & 3) sind, sind die dominanten Schallquellen die Gegenphase-Komponenten. Die virtuelle Schallquelle scheint sich in die entgegengesetzte Richtung zu bewegen als vorgesehen. Die Folgen davon sind schwerwiegend, wenn die Schallquelle mit einer Videoquelle kombiniert wird in einer Eintauch-Umgebung. Dass sich der Schall und das Bild in entgegengesetzte Richtungen bewegen, ist eine offensichtlich nicht akzeptable Form eines modalen Konflikts.

- B'-Format

Die beobachteten Effekte sind dieselben wie für das B-Format-Signal.

- B-Format mit Decoder-Verzerrung

Eine klare Pfad-Trajektorie, wenngleich nicht-linear, aufgrund des Entferns der Gegenphase-Komponenten.

- B'-Format mit Decoder-Verzerrung

Eine klare lineare Trajektorie von dem vorne-rechts-Lautsprecher zu dem vorne-links-Lautsprecher.

## Patentansprüche

1. Verfahren zur Erzeugung eines Schallfelds aus einem Array von Lautsprechern, wobei das Array einen Hörraum definiert, in dem die Ausgaben der Lautsprecher kombiniert werden, um eine räumliche Wahrnehmung einer virtuellen Schallquelle zu liefern, wobei das Verfahren die Erzeugung einer jeweiligen Ausgabe-komponente  $P_n$  für jeden Lautsprecher in dem Array zum Steuern der Ausgabe des jeweiligen Lautsprechers aufweist, wobei die Ausgabe aus Daten abgeleitet wird, die in einem Eingangssignal getragen werden, wobei die Daten ein Summenreferenzsignal  $W$  und Richtungsschallkomponenten  $X$ ,  $Y$ , ( $Z$ ) aufweisen, welche die Schallkomponente in unterschiedlichen Richtungen, wie durch die virtuelle Schallquelle erzeugt, repräsentieren, wobei das Verfahren die Schritte aufweist des Erkennens für jeden Lautsprecher, ob die jeweilige Komponente  $P_n$  sich in Phase oder Gegenphase zu dem Summenreferenzsignal  $W$  ändert, des Modifizierens des Signals, wenn es sich in Gegenphase befindet, und des Zuführens der resultierenden modifizierten Komponenten zu den jeweiligen Lautsprechern.

2. Verfahren gemäß Anspruch 1, wobei die Richtungsschallkomponenten jeweils mit einem Verzerrungsfaktor multipliziert werden, der eine Funktion der jeweiligen Richtungsschallkomponente ist, so dass eine sich bewegende virtuelle Schallquelle, die einer gleichmäßigen Trajektorie folgt, wie von einem Zuhörer an einem Punkt in dem Hörfeld wahrgenommen, auch einer gleichmäßigen Trajektorie folgt, wie an einem anderen Punkt

in dem Hörfeld wahrgenommen.

3. Verfahren gemäß Anspruch 2, wobei der Verzerrungsfaktor eine quadratische oder höhere geradzahlige Potenz der Richtungskomponente ist.

4. Verfahren gemäß Anspruch 2, wobei der Verzerrungsfaktor eine Sinusfunktion der Richtungskomponente ist.

5. Vorrichtung zur Erzeugung eines Schallfelds, die aufweist ein Array von Lautsprechern, die einen Hörraum definieren, in dem die Ausgaben der Lautsprecher kombiniert werden, um eine räumliche Wahrnehmung einer virtuellen Schallquelle zu liefern, Mittel zum Empfang und zur Verarbeitung von Daten, die in einem Eingangssignal getragen werden, wobei die Daten ein Summenreferenzsignal  $W$  und Richtungsschallkomponenten  $X$ ,  $Y$ , ( $Z$ ) aufweisen, welche den Schall in unterschiedlichen Richtungen, wie durch die virtuelle Schallquelle erzeugt, anzeigen, Mittel zur Erzeugung einer jeweiligen Ausgabekomponente  $P_n$  zur Steuerung der Ausgabe jedes Lautsprechers in dem Array aus diesen Daten, Mittel zur Erkennung für jeden Lautsprecher, ob die jeweilige Komponente  $P_n$  sich in Phase oder Gegenphase zu dem Summenreferenzsignal  $W$  ändert, Mittel zum Modifizieren des Signals, wenn es sich in Gegenphase befindet, und Mittel zum Zuführen der resultierenden modifizierten Komponenten zu den jeweiligen Lautsprechern.

6. Vorrichtung gemäß Anspruch 5, die weiter aufweist Mittel zum Multiplizieren jeder Richtungskomponente mit einem Verzerrungsfaktor, der eine Funktion der jeweiligen Richtungsschallkomponente ist, so dass eine sich bewegende virtuelle Schallquelle, die einer gleichmäßigen Trajektorie folgt, wie von einem Zuhörer an einem Punkt in dem Hörfeld wahrgenommen, auch einer gleichmäßigen Trajektorie folgt, wie an einem anderen Punkt in dem Hörfeld wahrgenommen.

7. Vorrichtung gemäß Anspruch 6, wobei der Verzerrungsfaktor eine quadratische oder höhere geradzahlige Potenz der Richtungskomponente ist.

8. Vorrichtung gemäß Anspruch 6, wobei der Verzerrungsfaktor eine Sinusfunktion der Richtungskomponente ist.

Es folgen 12 Blatt Zeichnungen

Fig.1.

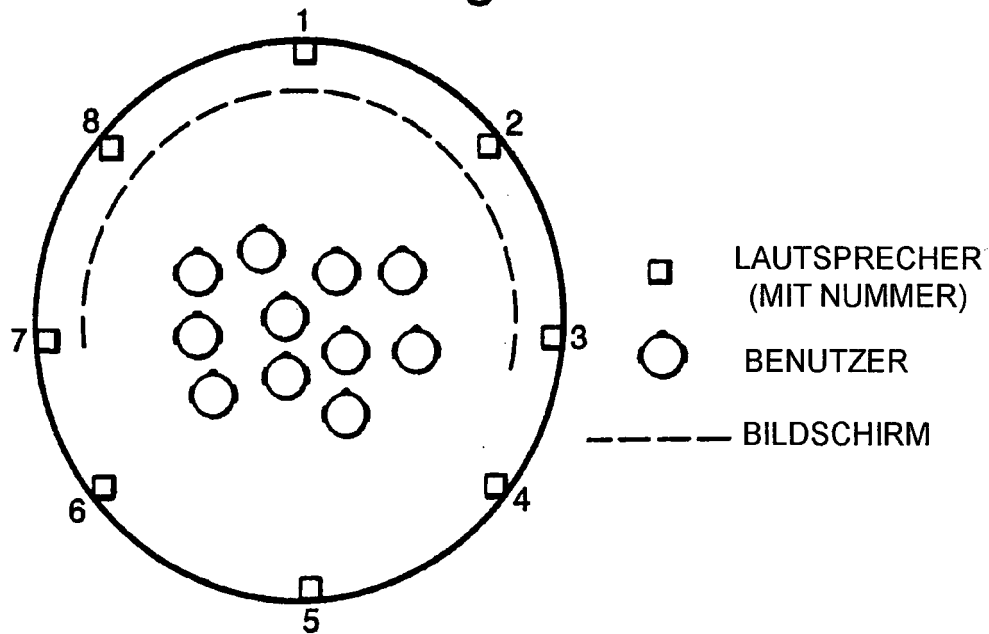


Fig.2.

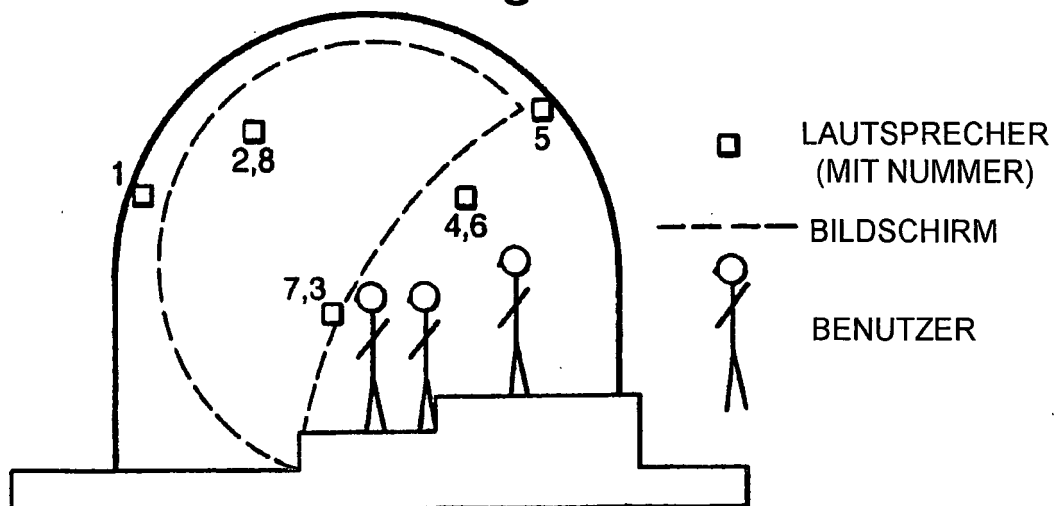


Fig.3.

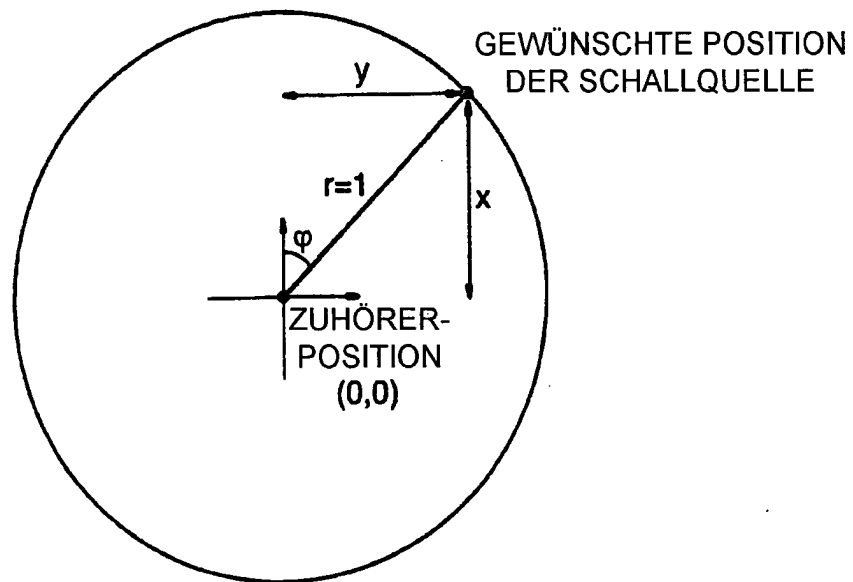
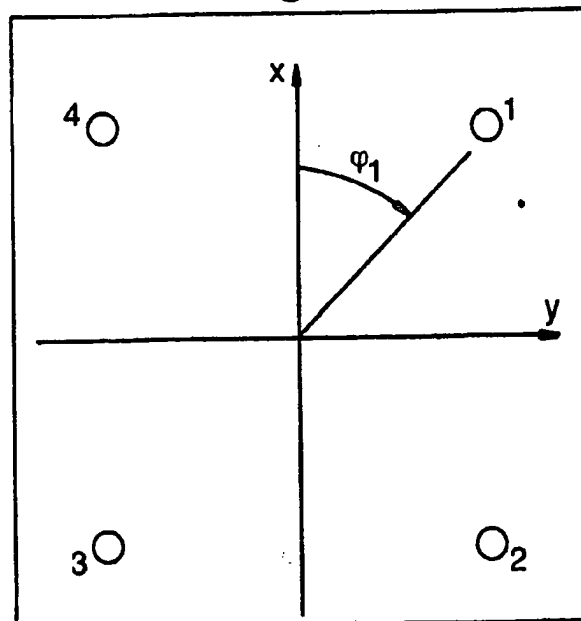
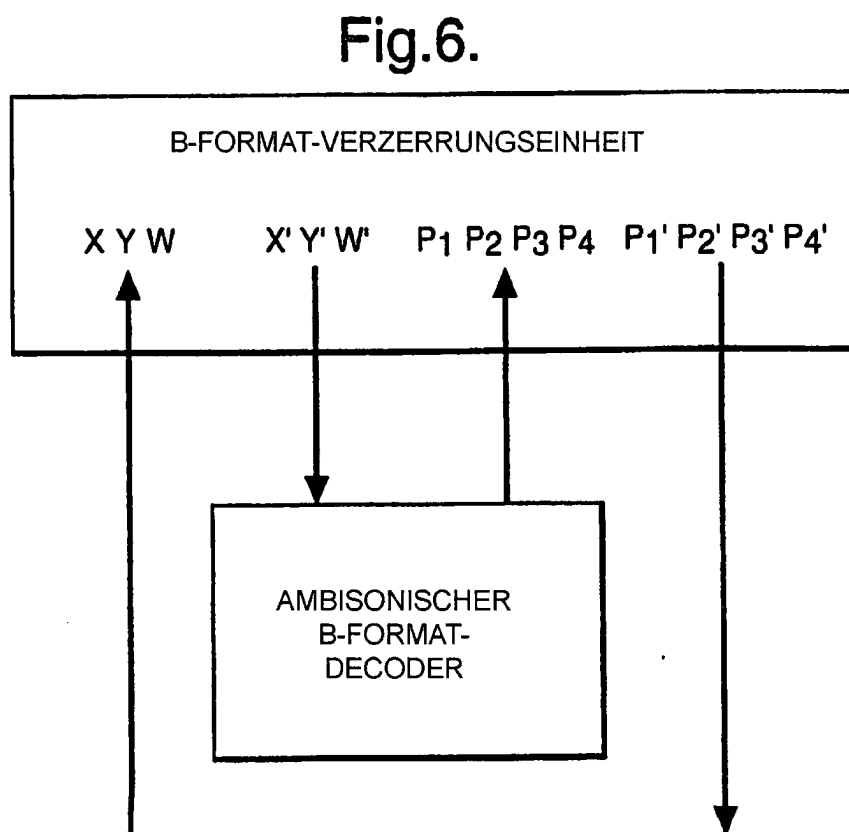
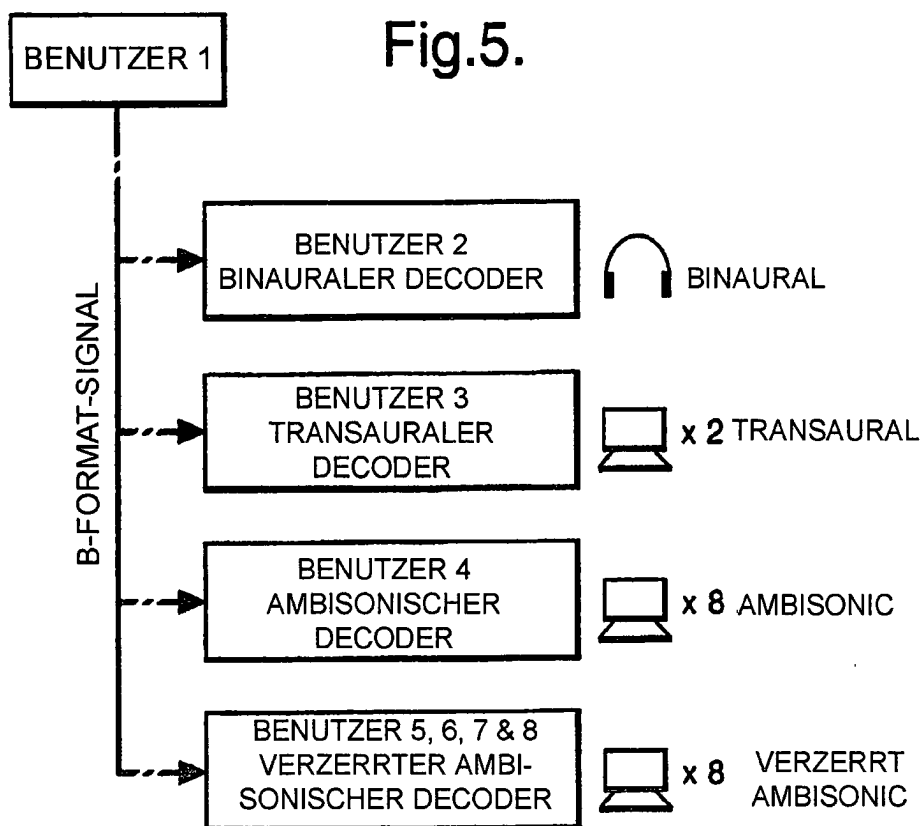


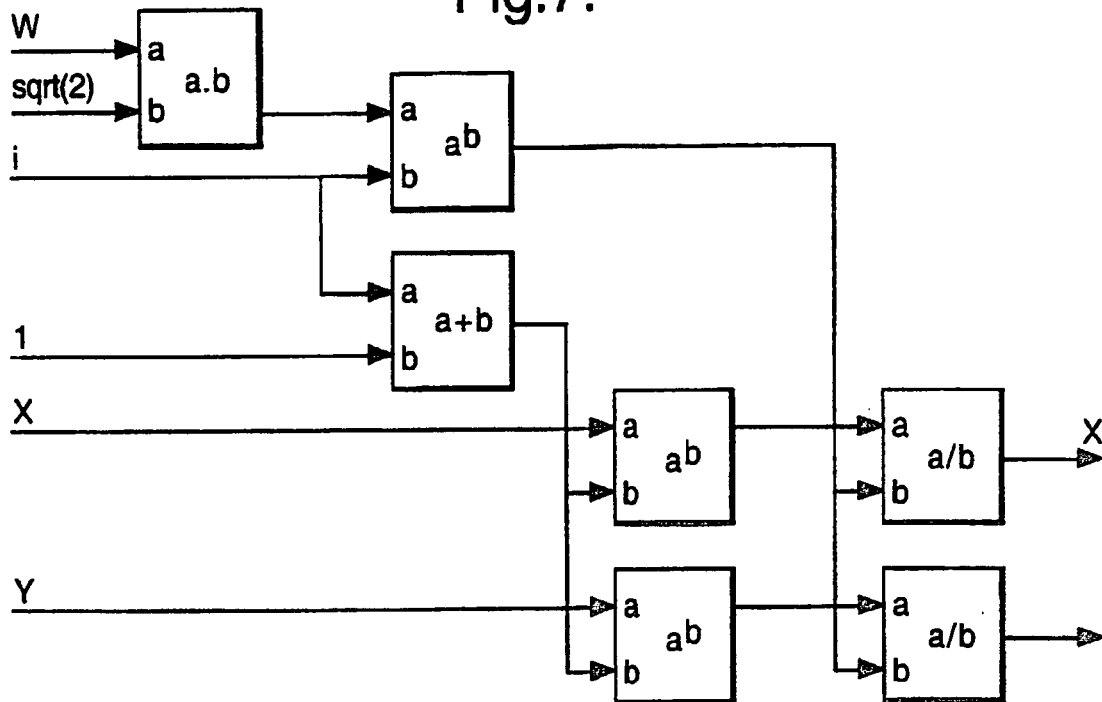
Fig.4.







**Fig.7.**



**Fig.8.**

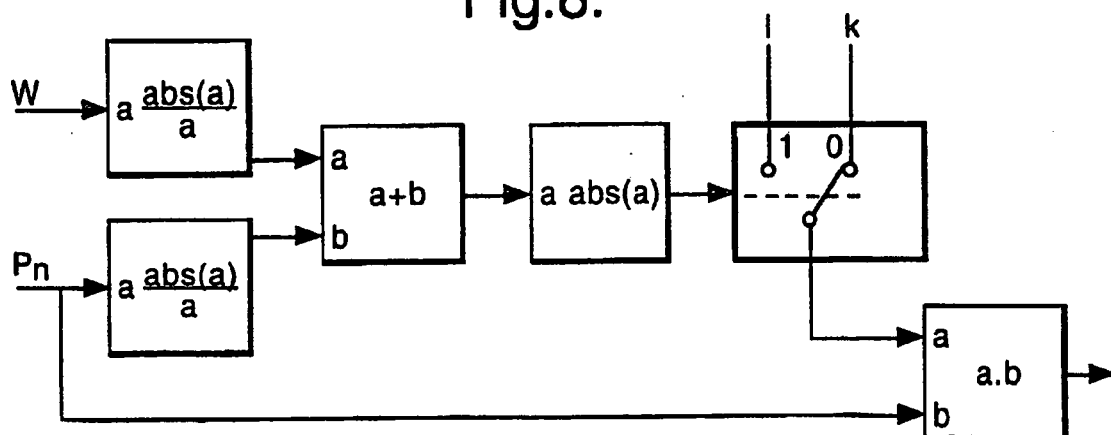


Fig.9.

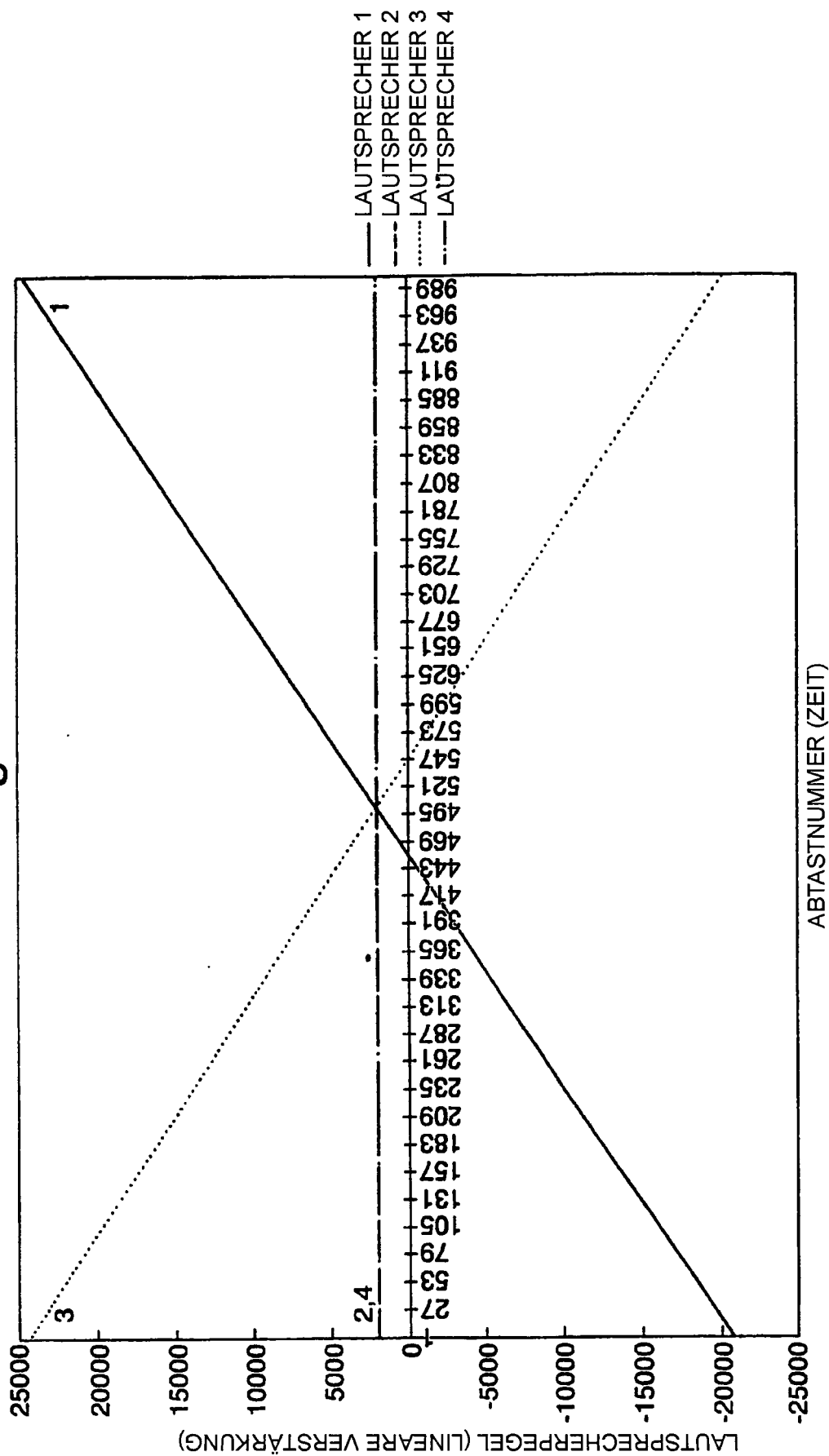


Fig.10.

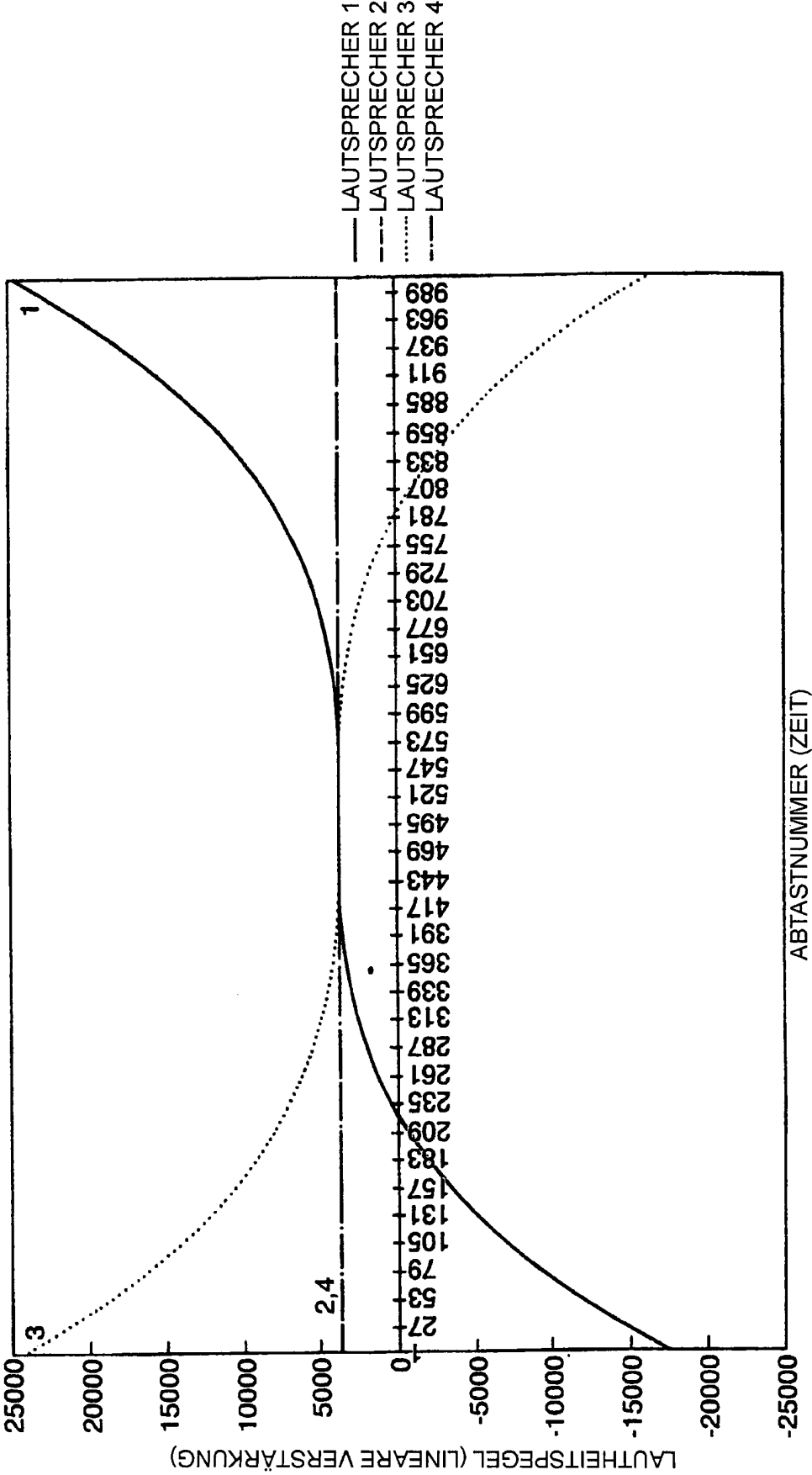
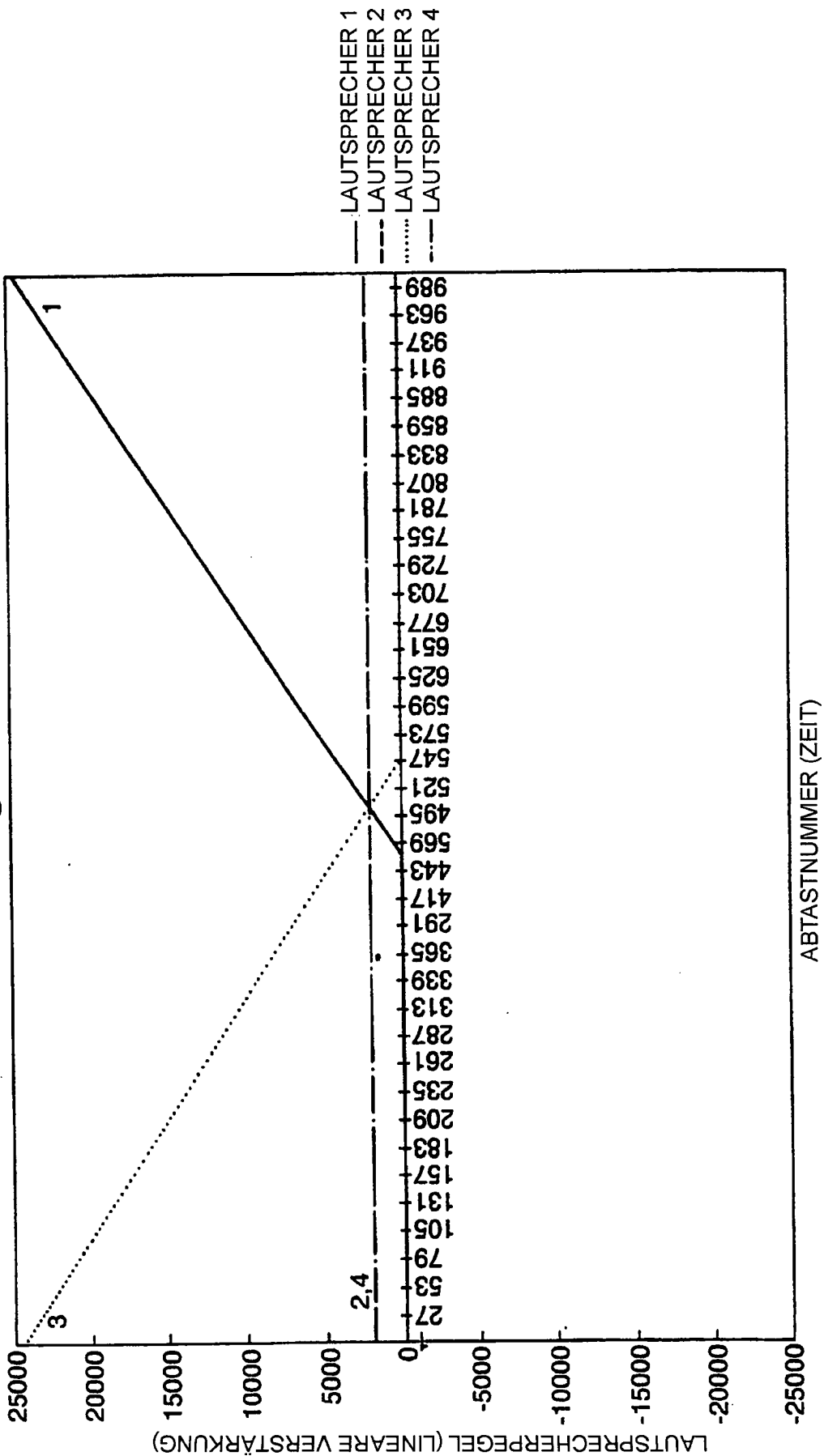


Fig.11.



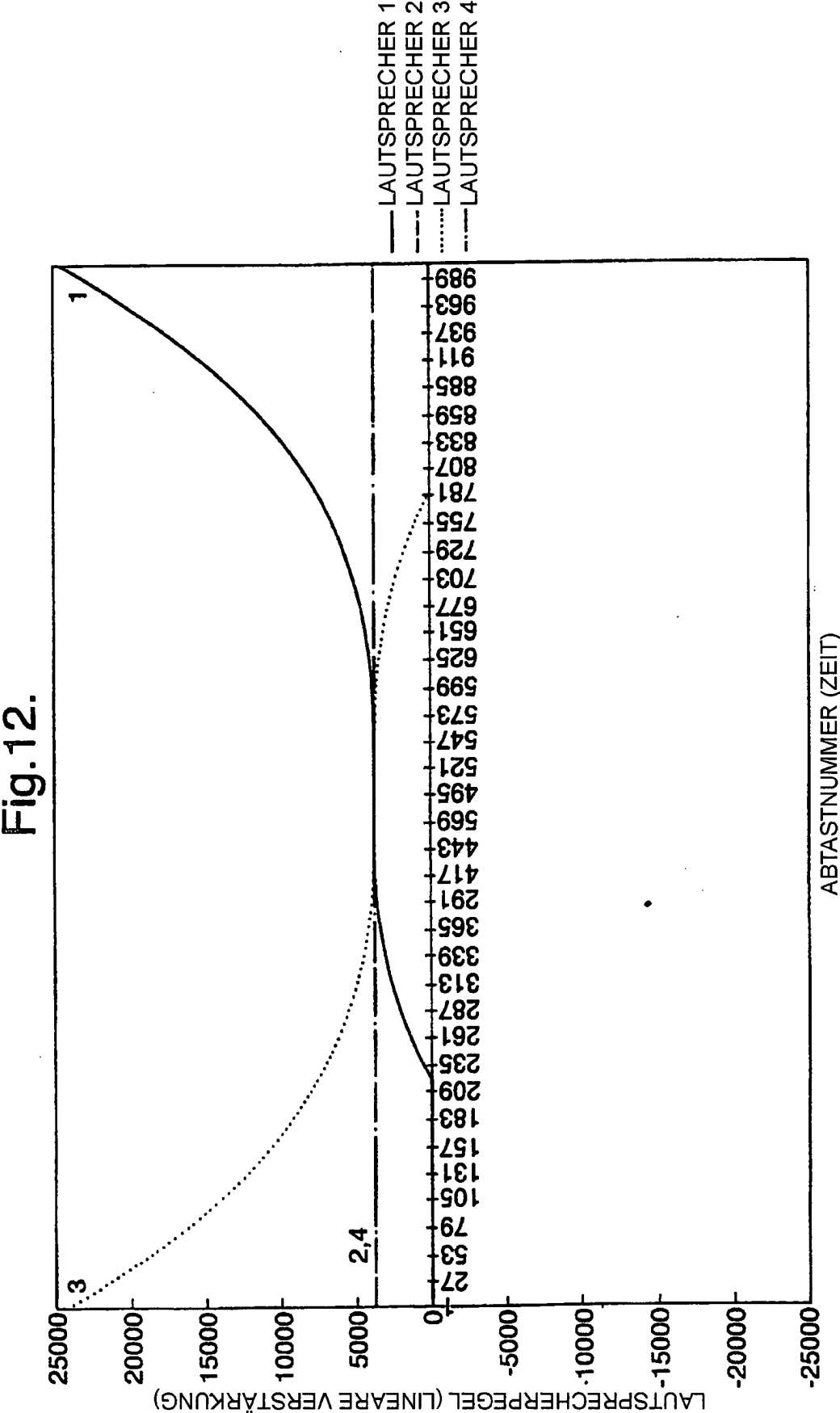


Fig.13.

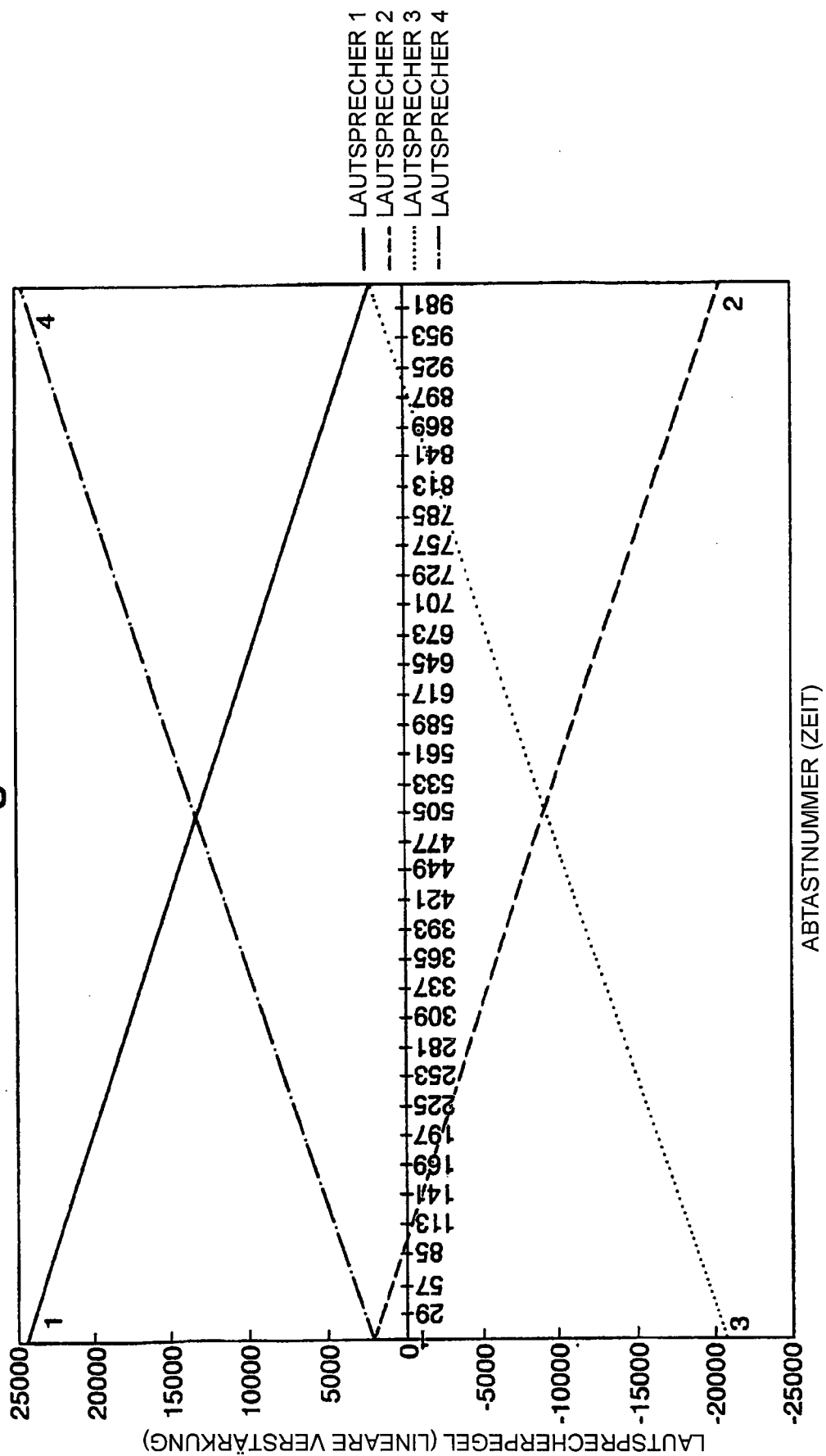


Fig.14.

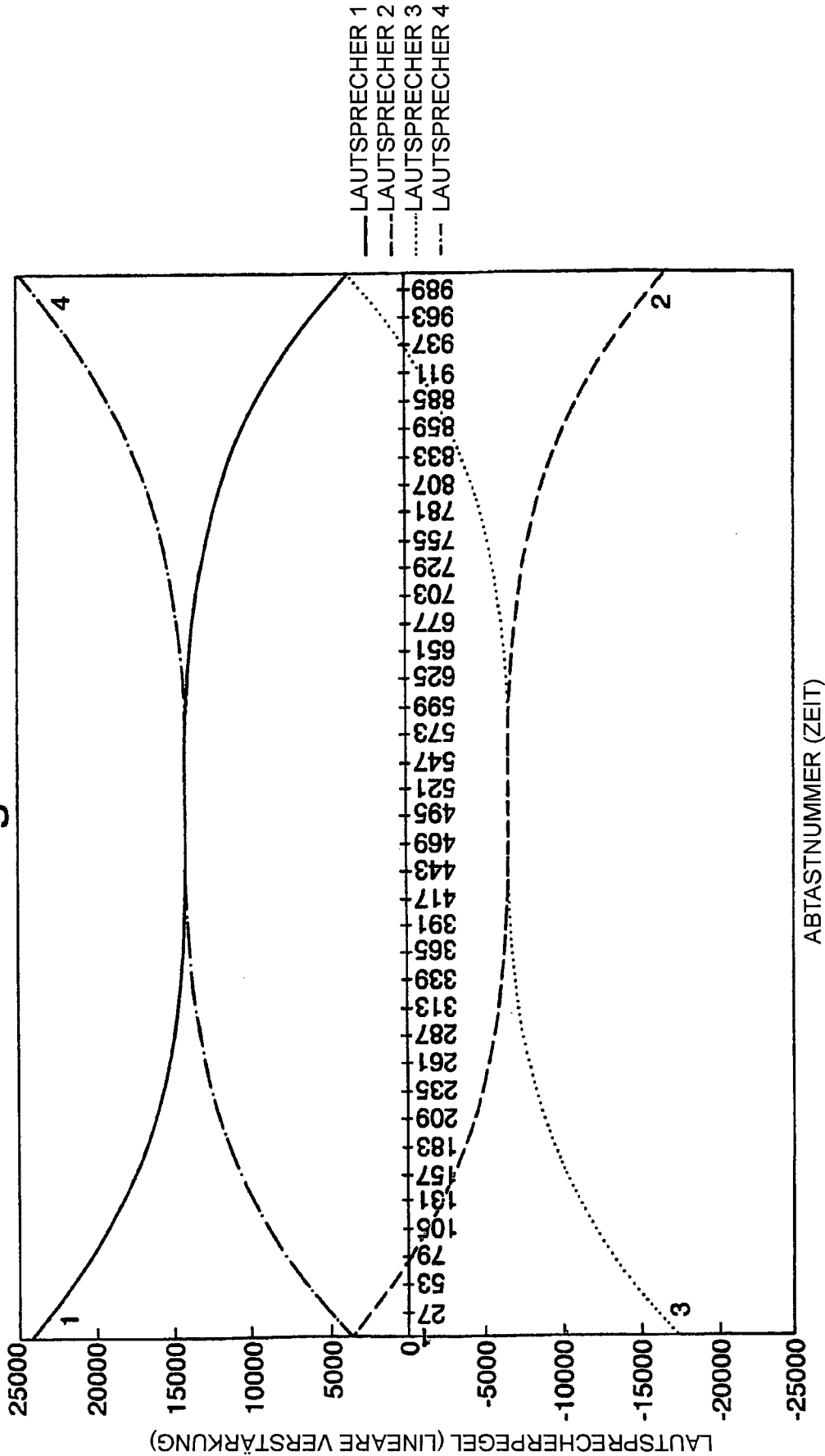




Fig.15.

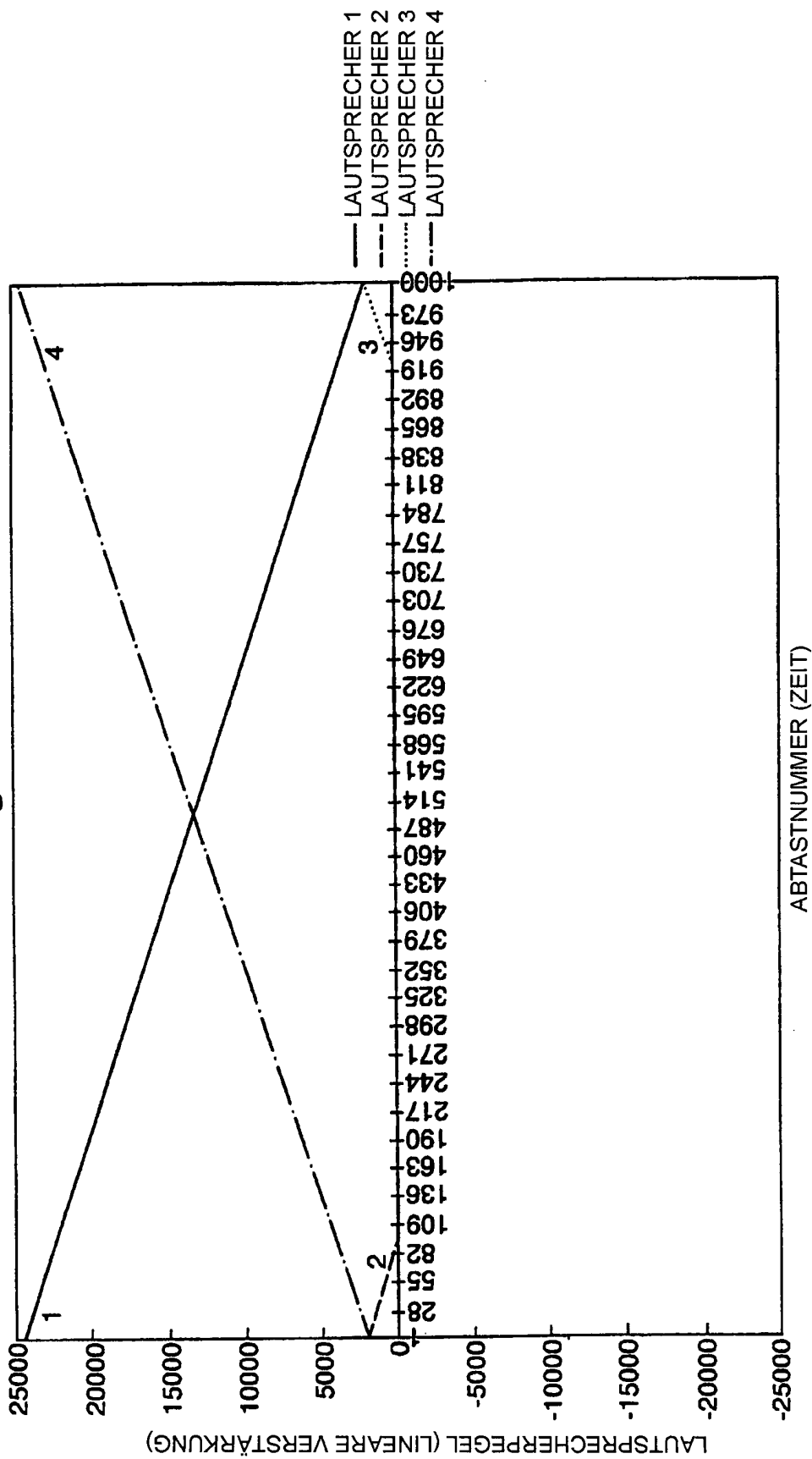


Fig.16.

