(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
15. Juli 2010 (15.07.2010)

PCT

(10) Internationale Veröffentlichungsnummer
WO 2010/078938 A2

(51) Internationale Patentklassifikation:
G10L 21/02 (2006.01) *H04R 25/00* (2006.01)
G10L 13/02 (2006.01)

(21) Internationales Aktenzeichen: PCT/EP2009/009129

(22) Internationales Anmeldedatum:
18. Dezember 2009 (18.12.2009)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2008 062 775.5
18. Dezember 2008 (18.12.2008) DE
10 2008 062 776.3
18. Dezember 2008 (18.12.2008) DE
10 2008 063 279.1
29. Dezember 2008 (29.12.2008) DE
10 2008 063 367.4
30. Dezember 2008 (30.12.2008) DE
10 2009 018 470.8
22. April 2009 (22.04.2009) DE
10 2009 018 469.4
22. April 2009 (22.04.2009) DE
10 2009 032 236.1 8. Juli 2009 (08.07.2009) DE
10 2009 032 238.8 8. Juli 2009 (08.07.2009) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **FORSCHUNGSGESELLSCHAFT FÜR ARBEITSPHYSIOLOGIE UND ARBEITSSCHUTZ E. V.** [DE/DE]; Ardeystr. 67, 44139 Dortmund (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder *(nur für US)*: **BAUER, Hans-Dieter** [DE/DE]; Gäuseland 16, 44227 Dortmund (DE). **PLINGE, Axel** [DE/DE]; Annenstr. 5, 44137 Dortmund (DE).

(74) Anwalt: **SCHNEIDERS & BEHRENDT**; Huestrasse 23, 44787 Bochum (DE).

(81) Bestimmungsstaaten *(soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD AND DEVICE FOR PROCESSING ACOUSTIC VOICE SIGNALS

(54) Bezeichnung : VERFAHREN UND VORRICHTUNG ZUM VERARBEITEN VON AKUSTISCHEN SPRACHSIGNALEN

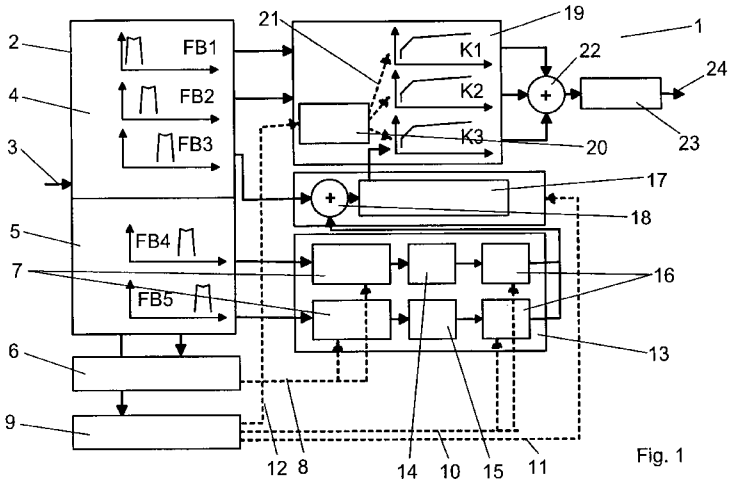


Fig. 1

(57) Abstract: The invention relates to a method for processing acoustic voice signals using an electronic processing device. According to the invention, in order to provide an improved processing of acoustic voice signals relative to the state of the art, a processing according to the type of sound is carried out, wherein poorly articulated sounds are lengthened out. The invention further relates to a method for processing acoustic voice signals in which poorly articulated sounds are quickly recognised and replaced by corresponding synthesised sounds. A further aspect of the invention concerns a speech synthesis method, wherein two or more format wave forms are generated by modulation of a source signal oscillating at a format frequency with an envelope function, the two or more format waveforms are added and the added format waveforms are concatenated According to a pitch interval length and according to suprasegmental concatenation rules to give a suprasegmental voice signal. According to the invention, the source signals on generating the format waveforms are frequency modulated. The invention finally related to a method for controlling the matching of a hearing device by using test signals with natural or near natural speech elements which are spectrally filtered or selected such that the spectrum of the test signal corresponds to the spectral range of at least one filter of a filter bank in the hearing device.

(57) Zusammenfassung:

[Fortsetzung auf der nächsten Seite]

WO 2010/078938 A2

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Bestimmungsstaaten *(soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— *ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts (Regel 48 Absatz 2 Buchstabe g)*

Die vorliegende Erfindung betrifft ein Verfahren zum Verarbeiten von akustischen Sprachsignalen unter Verwendung einer elektronischen Verarbeitungseinrichtung. Um eine gegenüber dem Stand der Technik verbesserte Verarbeitung von akustischen Sprachsignalen zur Verfügung zu stellen, wird mit der Erfindung vorgeschlagen, dass eine lautklassenspezifische Verarbeitung der Sprachsignale erfolgt, wobei schwach artikulierte Laute zeitlich verlängert werden. Weiterhin betrifft die Erfindung ein Verfahren zum Verarbeiten von akustischen Sprachsignalen, bei dem schwach artikulierte Laute in kurzer Zeit erkannt und durch ihnen entsprechende, synthetische Laute ersetzt werden. Ein weiterer Aspekt der Erfindung betrifft ein Sprachsyntheseverfahren, wobei zwei oder mehr Formantwellenformen jeweils durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit einer Umhüllenden-Funktion erzeugt werden, die zwei oder mehr Formantwellenformen addiert werden und die addierten Formantwellenformen gemäß einer Pitch-Intervalllänge und gemäß suprasegmentaler Verkettungsregeln zu einem suprasegmentalen Sprachsignal verkettet werden. Gemäß der Erfindung werden die Quellensignale bei der Erzeugung der Formantwellenformen frequenzmoduliert werden. Schließlich betrifft die Erfindung ein Verfahren zur Kontrolle der Anpassung eines Hörgerätes, wobei Testsignale verwendet werden, die natürliche oder naturähnliche Sprachelemente sind, die in der Weise spektral gefiltert oder ausgewählt werden, dass das Spektrum des Testsignals zu dem Spektralbereich wenigstens eines Filters einer Filterbank des Hörgerätes korrespondiert.

Verfahren und Vorrichtung zum Verarbeiten von akustischen Sprachsignalen

Die vorliegende Erfindung betrifft ein Verfahren zum Verarbeiten von akustischen Sprachsignalen sowie eine hierzu geeignete Vorrichtung.

Entsprechende Verfahren und Vorrichtungen finden beispielsweise Anwendung in der Hörgerätetechnik, um für Personen mit Hörschäden eine Verbesserung der Verständlichkeit menschlicher Sprache herbeizuführen. Derartige herkömmliche elektro-akustische Anlagen weisen in der Regel Anordnungen von linear verstärkenden Baugruppen auf. Eine solche Baugruppe kann beispielsweise ein Mikrophoneingang, eine Filterbank, ein Kompressor oder ein Ausgangsverstärker sein. Dabei werden die akustische Sprachsignale zunächst über ein Mikrofon in elektrische Sprachsignale umgewandelt, welche in den Mikrophoneingang eingegeben werden. Mittels der Filterbank, welche meist mehrere Bandpassfilter aufweist, erfolgt eine Aufteilung des elektrischen Sprachsignals in mehrere Frequenzbänder, welche von dem Kompressor individuell komprimiert werden, wozu dieser mehrere Kompressoruntereinheiten aufweist. Anschließend werden die komprimierten Frequenzbänder zu einem komprimierten Sprachsignal miteinander kombiniert, welches mittels des Ausgangsverstärkers verstärkt wird.

Bei ungestörter Hörumgebung oder ungestörtem Hörvermögen besitzen Sprachelemente starke Redundanzen. Deren physiologische Wirkungen gehen bei sensorischen Hördefiziten und auch bei zunehmenden Pegeln von Umgebungslärm verloren. Speziell sensorische Hördefizite erzeugen Erhöhungen der Hörschwelle, die mit Recruitment und damit mit Einengungen

der nutzbaren Hördynamikweite verbunden sind. Sie fallen besonders stark bei den häufig vorkommenden Hochtonverlusten aus, bei denen der für die Übertragung sehr wichtige Bereich der hoch liegenden, zweiten Formanten zum Teil sehr stark eingeschränkt wird, was schwerwiegende Folgen für die Sprachverständlichkeit hat.

Bei herkömmlichen Hörgeräten ist es meist nicht möglich, artikulativ schwache Laute mit ausreichender Stärke in das cortikale Restgehör oder die cortikale Restdynamik zu übertragen, weil der obere Hörbereich begrenzt ist oder weil die hierfür erforderliche Verstärkung nicht erreicht werden kann. Letzteres tritt häufig wegen einer einsetzenden akustischen Rückkopplung bei noch nicht ausreichender Verstärkung der schwach artikulierten Laute ein. Umgebungslärm wirkt dann besonders erschwerend und kann die Sprachverständlichkeit zusätzlich stark reduzieren. Bei starken Hörverlusten ist der hiervon Betroffene im Grenzfall gezwungen, auf eine akustische Sprachkommunikation zu verzichten und auf Lippenlesen auszuweichen. Dies trifft auch zu bei Personen, die nur Hörverluste geringen Grades aufweisen, und zwar dann, wenn die Umgebung mehr oder weniger lärmerfüllt ist. Dies ist zum Beispiel der Fall bei Ansagen in laufenden Fahrzeugen, auf Bahnhöfen oder in der „round-table" Diskussions-Situation.

Außerdem kann festgestellt werden, dass bei üblichen Hörgerätekonstruktionen der zur Verfügung stehende, pegelgesteuerte Kompressor das Sprachsignal deutlich verschlechtert, wenn der verfügbare Dynamikbereich schmaler wird als der Abstand der Pegel von schwachen und starken Lauten. Wenn dann die artikulativ schwachen Laute gut überschwellig gemacht werden, führt dies dazu, dass die artikulativ starken Laute in den Begrenzungskennlinienast des Kompressors hineingesteuert werden, was zu einer Verzerrung des Rhythmus und der Co-Modulation dieser Laute führt.

Ausgehend von diesem Stand der Technik ist es die Aufgabe der Erfindung, ein verbessertes Verfahren sowie eine verbesserte Vorrichtung zum Verarbeiten von akustischen Sprachsignalen bereitzustellen.

Diese Aufgabe wird bei einem Verfahren der eingangs genannten Art dadurch gelöst, dass eine lautklassenspezifische Verarbeitung der Sprachsignale erfolgt, wobei schwach artikulierte Laute zeitlich verlängert werden. Das kann dadurch erfolgen, dass die Energie schwacher Laute durch zeitversetzte Wiederholung eines merkmalstragenden Teils der Laut-Wellenform gestärkt wird.

Eine Lautklasse umfasst erfindungsgemäß alle klanglichen Variationen eines Lautes, welche sich von einem anderen Laut unterscheiden lassen. Beispielsweise kann ein „i" hoch, tief oder lang ausgesprochen werden, ohne die Grenzen der Lautklasse „i" zu verlassen.

Erfindungsgemäß erfolgt eine Be-, Ver- und/oder Aufarbeitung von schwach artikulierten Lauten der menschlichen Sprache selektiv oder in anderer spezifischer Weise. Durch diese individuelle Verarbeitung von Lauten kommt es nicht zu einer einfachen und für alle Sprachmerkmale einheitlichen Pegelsteuerung, welche mit den oben beschriebenen Nachteilen verbunden ist. Auch bei größerem Umgebungslärm kann eine Verarbeitung der akustischen Sprachsignale derart erfolgen, dass eine sehr gute Sprachverständlichkeit gegeben ist. Zur Verbesserung der Sprachverständlichkeit werden erfindungsgemäß schwach artikulierte Laute zeitlich verlängert, indem die schwachen Laute durch zeitversetzte Wiederholung eines Teils der Laut-Wellenform gestärkt werden. Da nicht wie beim Stand der Technik eine Verdeutlichung der schwach artikulierten Laute nur mittels einer ausreichenden Verstärkung derselben erzeugt wird, kommt es nicht zu einer störenden, verstärkungsbedingten Schwingneigung. Die Verarbeitung der akustischen Sprachsignale kann insgesamt sehr viel genauer an den jeweils vorliegenden Hörschaden angepasst werden.

Nach einer vorteilhaften Ausgestaltung der Erfindung werden die Sprachsignale in mehrere Frequenzbänder aufgeteilt. Dieses ermöglicht neben der lautklassenspezifischen Verarbeitung der Sprachsignale eine weitere Möglichkeit der individuellen Verarbeitung der Sprachsignale, so dass die Verarbeitung auch hierdurch an das jeweils vorliegende Hördefizit höchst genau angepasst werden kann.

Vorzugsweise erfolgt eine Aufspaltung der Sprachsignale in hochfrequente Frequenzbänder, die oberhalb einer oberen Grenzfrequenz liegen, und Frequenzbändern, die unterhalb der oberen Grenzfrequenz liegen. Die Grenzfrequenz entspricht vorzugsweise dem oberen Rand des Hörbereichs und kann individuell an das Ausmaß des jeweils vorliegenden Hochtonverlustes angepasst werden.

Mit der Erfindung wird weiter vorgeschlagen, dass die hochfrequenten Frequenzbänder zu niedrigeren Frequenzen unterhalb der oberen Grenzfrequenz und oberhalb einer unteren Grenzfrequenz verschoben werden. Hierdurch können Laute, die am oberen Rand des Hörbereichs oder jenseits der Grenze der Hörbarkeit, liegen, spektral in einen besser nutzbaren niederfrequenten Hörbereich verschoben werden, so dass die Wirksamkeit dieser Laute gesteigert wird. Die Verschiebung der hochfrequenten Frequenzbänder zu niedrigeren Frequenzen unterhalb der oberen Grenzfrequenz muss jedoch die physiologische Klassenbildung der Sprachlaute vollständig in Takt belassen. Die Verschiebung darf also nur soweit oder nur auf eine solche Art geschehen, dass die natürlichen Klassengrenzen, die im physiologischen Klassifizierungsraum natürlicherweise zu finden sind, nicht überschritten werden. Insbesondere sind Inter-Laut-Transformationen auszuschließen. Beispielsweise darf durch die Frequenzverschiebung nicht ein „i“ zu einem „ü“ werden. Die Frequenzverschiebung darf nur in Form von Intra-Laut-Transformationen erfolgen, bei denen keine Umwandlung von Lauten erfolgt und bei denen beispielsweise aus einem hoch und spitz wahrnehmbaren „i“ ein dumpf wahrnehmbares „i“ wird. Um Überschreitungen von physiologischen Laut-Klassengrenzen zu verhindern, erfolgt des Weiteren die Verschiebung der hochfrequenten Frequenzbänder zu niedrigeren Frequenzen oberhalb einer unteren Grenzfrequenz.

Eine weitere vorteilhafte Ausgestaltung der Erfindung sieht vor, dass die Verschiebung der hochfrequenten Frequenzbänder individuell in Abhängigkeit der jeweiligen Frequenzlage des hochfrequenten Frequenzbandes erfolgt. Auch durch diese Ausgestaltung ist eine sehr individuelle Verarbeitung der akustischen Sprachsignale möglich, indem keine einheitliche Verschiebung von hochfrequenten Frequenzbändern erfolgt, welche beispielsweise auch zu den oben beschriebenen Inter-Laut-Transformationen führen kann, sondern die

Frequenzbänder werden individuell in Abhängigkeit ihrer jeweiligen Frequenzlage unter ausschließlich Intra-Laut-Transformationen in einen geeigneten Frequenzbereich verschoben.

Gemäß einer weiteren vorteilhaften Ausgestaltung der Erfindung werden die unterhalb der oberen Grenzfrequenz liegenden Frequenzbänder mit unterschiedlicher Pre-Emphasis versehen. Diese Ausgestaltung der Erfindung dient insbesondere der Verbesserung des Signal-Rausch-Verhältnisses. Da die unterhalb der oberen Grenzfrequenz liegenden einzelnen Frequenzbänder in unterschiedlichen Frequenzbereichen angeordnet sind, ist es sinnvoll, jedes dieser Frequenzbänder mit einer unterschiedlichen Pre-Emphasis zu modulieren. Auch diese Vorgehensweise kommt der individuellen Anpassbarkeit des Verfahrens an das jeweils vorliegende Hördefizit zugute.

Zweckmäßigerweise werden die unterhalb der oberen Grenzfrequenz liegenden Frequenzbänder unterschiedlich komprimiert. Auch hierdurch kann den jeweiligen Anforderungen an die Verarbeitung der akustischen Sprachsignale genüge getan werden, indem diese sehr individuell verarbeitet werden.

Es wird weiter vorgeschlagen, dass den Sprachsignalen (oder besser: den Sprachsignalabschnitten) jeweils eine bestimmte Lautklasse zugeordnet wird. Hierzu kann beispielsweise ein Lautklassenselektor eingesetzt werden, mit dem ein Abgleich der Sprachsignale mit vorgegebenen Charakteristika der einzelnen Lautklassen erfolgen kann, so dass feststellbar ist, zu welcher Lautklasse der mit dem jeweiligen Sprachsignal erhaltene Laut gehört. Vorzugsweise erfolgt auf Basis dieses Abgleichs eine individuelle Steuerung der einzelnen Verarbeitungsmaßnahmen der Sprachsignale gemäß der Erfindung.

Vorzugweise werden die zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbänder in Abhängigkeit der den Sprachsignalen jeweilig zugeordneten Lautklasse zu einem Zwischensprachsignal miteinander kombiniert. Ob und in welcher Form diese Kombination durchgeführt wird, kann ebenfalls individuell an die jeweiligen Anforderungen angepasst werden.

Es wird des Weiteren als vorteilhaft erachtet, wenn die zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbänder mit einem oberen Frequenzband, welches unterhalb der oberen Grenzfrequenz am nächsten zu dieser angeordnet ist, zu einem Hochtonzwischensprachsignal kombiniert werden. Durch diese Zuführung des oberen Spektralbandbereiches zu den zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbändern erfolgt eine Mischung von Signalenergien, die hörschadenspezifisch beispielsweise oberhalb von 2,3 KHz liegen, und den zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbändern, wodurch diese Mischung perzeptiv verstärkt wird, was einer Pegelanhebung von etwa 3 bis 6 dB entspricht. Diese neuartige Hochtonband-Mischung wird im Anschluss individuell komprimiert und mit den übrigen individuell komprimierten Frequenzbändern zu einem Ausgangssprachsignal kombiniert.

Es wird weiter vorgeschlagen, dass das Zwischensprachsignal bzw. das Hochtonzwischensprachsignal in Abhängigkeit von der den Sprachsignalen jeweils zugeordneten Lautklasse gespeichert, in vorgebbaren Zeitabständen abgerufen, individuell komprimiert und zur Erzeugung eines Ausgangssprachsignals mit den übrigen individuell komprimierten Frequenzbändern, die unterhalb der oberen Grenzfrequenz liegen, kombiniert wird. Bei stimmhafter Sprache ist die natürliche Dämpfung der oberen Formantresonanzen so stark, dass die Umhüllenden relativ schmale Spitzen und breite Täler aufweisen. In diesem Fall können durch die gemäß dieser Ausgestaltung der Erfindung erfolgende Wiederholung der jeweiligen Wellenform mit einer zeitlichen Verzögerung die Täler störungsfrei aufgefüllt werden, wodurch der Formant-Energieinhalt der Gesamtschwingung wesentlich, beispielsweise bis zu 6 dB, erhöht wird. Bei intakter Energiesummation des Ohres über Segmente von 10 ms kann hierdurch ein erheblicher physiologischer Wirkungszuwachs beispielsweise bezüglich der Lautheit und der Klarheit erzeugt werden. Ein schwach artikulierter Laut wird durch diese Verarbeitungsmaßnahmen zeitlich verlängert.

Gemäß einer weiteren vorteilhaften Ausgestaltung der Erfindung wird das Ausgangssprachsignal zur Kompensation von ungewünschten spektralen Eigenschaften von an die Verarbeitungseinrichtung anschließbaren Sprach-

signalausgabeeinheiten mittels eines Equalizers moduliert. Vorzugsweise weist der Equalizer ein programmierbares FIR-Filter auf.

Nach einer zweckmäßigen Ausgestaltung der Erfindung erfolgt eine Vorverstärkung der Frequenzbänder vor ihrer Kompression in Abhängigkeit von der den Sprachsignalen jeweils zugeordneten Lautklasse und/oder der Lautstärke von Umgebungsgeräuschen. Bei stärkeren Umgebungsgeräuschen wird die Verstärkung pro Band adaptiv wahlweise so gesenkt, dass ein mittlerer Pegel entsteht, der individuell entweder gerade gut wahrnehmbar oder gerade nicht mehr wahrnehmbar ist. Je nach Situation kann der schwerhörende Mensch die Voreinstellung der gerade guten, permanenten Wahrnehmbarkeit des Umgebungslärms wählen, wenn eine Kontrollmöglichkeit bestehen bleiben soll, oder die Voreinstellung „gerade nicht mehr wahrnehmbar", wenn jeglicher Umgebungslärm als störend aufzufassen ist.

Nach einer vorteilhaften Ausgestaltung der Erfindung umfasst die lautklassenspezifische Verarbeitung der Sprachsignale für jede Lautklasse eigene oder wenigstens zwei lautklassenübergreifende Verarbeitungsmaßnahmen. Bei den lautklassenspezifischen, wenigstens zwei lautklassenübergreifenden Verarbeitungsmaßnahmen sind insbesondere solche zu wählen, die für mehrere Lautklassen gleichermaßen anwendbar sind und einen Wahrnehmungsgewinn ohne Störungen erzeugen.

Ferner wird es als vorteilhaft erachtet, wenn das niedrigste, zu niedrigen Frequenzen verschobene, hochfrequente Frequenzband immer Signale liefert, wohingegen die frequenzmäßig höher liegenden, zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbänder lautklassenspezifisch zugeschaltet werden.

Mit dem erfindungsgemäßen Verfahren erfolgt somit eine nichtlineare Zeitbereichsmodifikation und eine nichtlineare Frequenzbereichsmodifikation der eingehenden akustischen Sprachsignale, wobei diese Frequenzbereichsmodifikationen eng aufeinander abgestimmt sind. Die Signalmodifikation im Zeitbereich in Form einer zeitlichen Verlängerung eines Sprachsignals erfolgt vorzugsweise nur bei einem spektral vorselektierten Teil des Sprachsignals,

insbesondere dort, wo eine solche Modifikation sinnvoll ist und keine Störungen hervorruft. Das betrifft insbesondere alle impulshaltigen Sprachanteile, die so kurz sind, dass sie von quasi Pausen ständig umgeben sind. Dies sind insbesondere Plosive und solche Teile von stimmhaft angeregten Lauten, deren höhere Formanten deutlich oberhalb von 1,2 KHz liegen. Bei diesen ist eine kompatible Verarbeitung möglich. Durch die erfindungsgemäße Auffüllung der breiten Täler der Umhüllenden der oberen Formantresonanzen wird der Formant-Energieinhalt der Gesamtschwingung wesentlich erhöht. Im Fall der Plosive kann man durch Speicherung, Verzögerung und Wiederaufaddieren bestimmter schwach artikulierter Laute den Explosionsbereich mit seiner Spitze, ohne Störungen oder Klassenänderungen zu erzeugen, mit der identischen Frequenzbereichsvorfilterung bis zu 10 ms ausdehnen und erreicht auch bei diesen Einmalereignissen durch die oben genannte Verzögerung und Summation einen deutlichen Wirkungszuwachs relativ zu den nicht impulshaltigen Merkmalssignalen. Da auch Formant-Übergänge von Vokalen, die bekanntermaßen auch für kontextgebundene Plosive merkmalstragend sind, nach dem beschriebenen Verfahren wesentlich mehr Energie erhalten, ist auch deren Wirkung in Verbindung mit vokalhaften Lauten, die hochliegende zweite Formanten etwa oberhalb von 1,3 KHz enthalten, verstärkt. Dadurch können kontextuelle Plosive in Verbindung mit den Lauten „ö", „ü", „i" und „e" an Klarheit oder Prägnanz gewinnen. Wesentlich für die Erfindung ist, dass die Hervorhebung von schwach artikulierten Lauten nicht durch Erhöhung der Verstärkung erfolgt und daher auch keine verstärkungsbedingte zusätzliche Schwingneigung hervorruft. Jedoch kann ein erfindungsgemäß hervorgehobener Laut auch verstärkt werden, wenn dieses zweckdienlich ist und nicht zu den im Zusammenhang mit dem Stand der Technik genannten Nachteilen führt.

Im unteren Frequenzbereich von 2,5 bis 4 KHz wird der zweite Formant des „i" durch Frequenzverschiebung des dritten Formanten um einen Faktor von ca. 0,8 und Überlagerung wesentlich robuster gemacht, also hervorgehoben. Da die Merkmalgrenze des „i" bei Lagen der zweiten Formanten bei 2,1 KHz gefunden wird, ist durch die Begrenzung der Frequenzverschiebung durch die untere Grenzfrequenz eine feste Beschneidung der niederfrequenten Energien bei 2,3 KHz durchzuführen. Die Merkmalsenergien der anderen frikativischen Laute werden kompatibel konzentriert und frequenzbegrenzt. Dies bewirkt vor allem

bei Menschen mit Hochtonhörverlusten eine bessere Wirkung der Merkmalsenergien bei „ch" und auch bei „f". Die untere Frequenzbegrenzung sorgt immer dafür, dass eine Anregung der physiologischen „sch"-Kanäle, also eine Inter-Laut-Klassenüberschreitung, unterbleibt.

Der obere Frequenzbereich von 5 bis 9 kHz enthält vor allem Merkmalsenergien des „s" aber auch des „t" und des „ch". Um in den Nutzhörbereich zu gelangen, ist eine andersgeartete mittlere Frequenzverschiebung um einen individuellen Faktor erforderlich. Hierdurch kann eine der natürlichen Schärfe-Wahrnehmung im hochliegenden Frequenzbereich äquivalente Wahrnehmung von Schärfe bei der verschobenen Merkmalsenergie und somit für eine physiologisch Natur-Laut-adäquate Wahrnehmung erfolgen. Die Frequenzverschiebungsfaktoren der einzelnen Frequenzbandverschiebeeinheiten werden bei Hörgerätanwendungen programmierbar gemacht, um Anpassungen an den individuellen Hörverlust zu ermöglichen. Die Vermeidung von Artefakten bei stimmtonangeregten Lauten, also insbesondere bei den genannten Vokalen einschließlich ihrer Formant-Übergänge, verlangt allerdings eine Synchronisierung der Rechen-Verarbeitungsfenster mit den Realzeit-Pitch-Perioden. Daher ist ein Pitch-Synchronisierer unverzichtbar. Bei Segmenten mit Frikativen kann die Merkmalsburstverlängerung keine Pausen auffüllen, da es keine gibt. Allerdings ist eine Überlagerung nach Verzögerung auch nicht schädlich. Eine spezielle Unterdrückung der Verzögerung ist daher nicht erforderlich.

Bei bekannten Verfahren zum Verarbeiten von akustischen Sprachsignalen in der Hörgerätetechnik können Sprachsignaltransformationen zur Kompensation des Hörverlustes eingesetzt werden, wobei ein transformiertes Sprachsignal mehr oder weniger direkt durch vergleichsweise einfache Modifikationen im Spektralbereich oder im Zeitbereich aus dem Originalsignal abgeleitet wird. Allerdings ist dies nur dann effektiv, wenn transformationsbegrenzende Randbedingungen eingehalten werden, die so gestaltet sind, dass Sprachlautverwechslungen durch inadäquate Überlagerung des Original-spektrums mit dem transformierten Spektrum verhindert werden.

Gemäß einem weiteren Aspekt der Erfindung werden schwach artikulierte Laute in dem zu verarbeitenden Sprachsignal in möglichst kurzer Zeit erkannt und durch ihnen entsprechende, synthetische Laute ersetzt.

Im Gegensatz zu der Anwendung von einfachen sprachsignalmodifizierenden Transformationen, wie sie aus dem Stand der Technik bekannt sind, erfolgt erfindungsgemäß ein selektiver Ersatz von Sprachsignalelementen, welche schwach artikuliert sind. Diese Selektivität wird mittels eines spezifisch auf den Zweck der Lautklassifizierung zugeschnittenen Sprachsignalerkennungsverfahrens erzeugt. Hierbei werden einzelne Lautklassen in kurzer Zeit selektiert. Bei Vorhandensein der schlecht wahrnehmbaren beziehungsweise schwach artikulierten Laute werden Ersatzlaute aus gespeicherten Bestandteilen synthetisiert und an Stelle der zu ersetzenden Laute in das ursprüngliche Sprachsignal eingefügt.

Anders als bei herkömmlichen Spracherkennungssystemen muss hier auf Lautebene eine Klassifikation in vergleichsweise kurzer Zeit vorgenommen werden, um die Wahrnehmung von Asynchronität zwischen Lippenbild und Sprachsignal beim Fremdsprecher und zwischen Sprachproduktion und Sprachaufnahme bei selbst artikulierter Sprache zu vermeiden. Bei digitalen Signalaufbereitungstechniken ist zu unterscheiden zwischen Verarbeitungszeiten und Prozessen für selbst produzierte Sprache einerseits und für Fremdsprecher-Sprache andererseits. Der erlaubte Zeitversatz, der gerade noch keine Wahrnehmung von Asynchronität hervorruft, liegt bei ca. 30 ms im Fall der Eigenartikulation und bei 80 ms im Fall der parallelen Aufnahme von Lippenbildern und Fremdsprecher-Sprachsignalen. Eine strikte Berücksichtigung dieser Randbedingungen ist erforderlich, wenn ein ungestörtes paralleles Lippenlesen als Informationsverstärkung erfolgen muss.

Die synthetischen Ersatzlaute oder deren Bestandteile werden vorzugsweise weitestgehend vorberechnet und in einem Speicher bevorratet. Vorab kann dafür gesorgt werden, dass diese neuen Laute den natürlichen Lauten wahrnehmungsmäßig weitestgehend ähnlich sind.

Nach einer vorteilhaften Ausgestaltung der Erfindung werden die synthetischen Laute vor ihrem Einfügen in das Sprachsignal hinsichtlich der Energie (Lautstärke) und/oder dem Frequenzschwerpunkt (Tonhöhe) an die zu ersetzenden, schwach artikulierten Laute angeglichen. Hierdurch werden die synthetischen Laute den zu ersetzenden, schwach artikulierten Lauten wahrnehmungsmäßig weitestgehend ähnlich.

Gemäß einer weiteren vorteilhaften Ausgestaltung der Erfindung wird das Sprachsignal vor dem Einfügen der synthetischen Laute zeitlich verzögert. Diese Verzögerung dient der zeitlichen Synchronisation von Sprachsignal und synthetischem Laut. Da die Verarbeitung des Sprachsignals, beispielsweise in Form einer Kompression, und die Sprachsignalerkennung sowie Erzeugung des synthetischen Lautes unterschiedliche Zeiten in Anspruch nehmen, ist die zeitliche Synchronisierung nahezu unerlässlich.

Eine weitere vorteilhafte Ausgestaltung der Erfindung sieht vor, dass die synthetischen Laute dynamisch in das empfangene Sprachsignal ein- und aus diesem ausgeblendet werden. Dieses bewirkt, dass störende und unnatürliche Sprünge im Klangbild vermieden werden.

In weiterer vorteilhafter Ausgestaltung der Erfindung wird das Sprachsignal in mehrere Frequenzbänder aufgeteilt, welche individuell komprimiert werden können, um eine ideale Anpassung der Sprachsignalverarbeitung an einen bestimmten Hörschaden zu ermöglichen.

Zweckmäßigerweise wird den schwach artikulierten Lauten jeweils ein vorgegebener Lautprototyp zugeordnet. Dies erfolgt durch Sprachsignal-erkennung. Vorzugsweise werden hierzu aus dem unbehandelten Sprachsignal solche Sprachsignalmerkmale extrahiert, die in Vorversuchen als für die Spracherkennung optimal geeignet identifiziert wurden. Zweckmäßigerweise erfolgt die Zuordnung der Lautprototypen zu den schwach artikulierten Lauten unter Berücksichtigung wenigstens eines Sprachsignalmerkmals.

Als solches Sprachsignalmerkmal können beispielsweise die spektralen Energieverhältnisse des Sprachsignals herangezogen werden. Hierzu können

beispielsweise 10 ms Energiemittelwerte in 4 Frequenzbändern ermittelt werden, welche in den folgenden Frequenzbändern liegen: E0 = 0,6 - 1,3 kHz, E1 = 1,2 - 2,4 kHz, E2 = 2,4 - 3,8 kHz und E3 = 4,5 - 8 kHz. Von diesen Bändern können die Verhältnisse, z.B. E1/E0, E2/E1, E3/E1 und E3/E2, gebildet und als Sprachsignalmerkmale verwendet werden.

Ein weiteres geeignetes Sprachsignalmerkmal kann die Stimmhaftigkeit des Sprachsignals sein. Zur Unterscheidung stimmhafter und stimmloser Laute wird beispielsweise der maximale Wert der normierten Kreuzkorrelation (NCC_{max}) über ein 10 ms Fenster herangezogen. Die normierte Kreuzkorrelation ist definiert als die Kreuzkorrelation (CC) zur Verschiebung t geteilt durch die Quadratwurzel des Produktes der Autokorrelation (AK) an den Stellen 0 und t. Das Maximum dieser Funktion im Bereich zwischen 1 und 10 ms wird als Indikator der Stimmhaftigkeit interpretiert.

$$NCC_{\max} = \max\{NCCF(t) / t \in 0...t_{\max}\}$$

$$\text{mit } NCCF(t) := \frac{CCF(t)}{\sqrt{(AKF(0)AKF(t)}}$$

Um Störeinflüsse zu vermeiden, sollte das Sprachsignal vorab von Gleichspannungs-Offsets befreit und tiefpassgefiltert werden. Vorzugsweise wird hierzu ein Tschebycheff-Tiefpass 4. Ordnung mit einer 3 kHz cutoff-Frequenz verwendet.

Als weiteres geeignetes Sprachsignalmerkmal kann eine Pause im Sprachsignal verwendet werden. Beispielsweise kann zur Erkennung von Verschlusspausen bei Plosiven ein „Peak-over-Average"-Pausendetektor eingesetzt werden. Dabei kann die lokale Modulation des Sprachsignals durch den Vergleich der absoluten Signalwerte (0,1 ms) auf einem langsameren (10 ms) Energiemittelwert ermittelt werden. Es können sowohl einkanalige Pausenerkenner wie auch mehrkanalige Pausenerkenner mit „min-max-Tracking" eingesetzt werden. Letztere sind weniger störanfällig als die einkanaligen Pausenerkenner. Zur Integration in den Pausenerkenner kann der letzte Pausenwert 20 ms gehalten werden.

Ein weiteres geeignetes Sprachsignalmerkmal ist die Anstiegsgeschwindigkeit (Rate-of-Rise; kurz ROR) des Sprachsignals. Die Anstiegsgeschwindigkeit kann beispielsweise zur Erkennung des Plosiv-Bursts verwendet werden, wobei es sich um den Zeitverlauf des Anstiegs lokal relativer Energie der Sprachsignal-Kurzzeit-Mittelwerte zu dem Zeitpunkt t ms und t+1 ms handelt. Als Maß kann das Verhältnis der Zeitmittelwerte über 20 ms zum Zeitpunkt t ms und t-1 ms gebildet werden. Zur zeitlichen Integration kann das Maximum dieses Wertes beispielsweise bei 50 ms gehalten werden. Zur Verbesserung der Erkennungsleistung bei t ms kann das Sprachsignal durch einen FIR-Bandpass mit einem Durchlassbereich von 2 - 10 kHz vorgefiltert werden.

$$ROR'(t) = \max\{ROR(t-h)/h = 0...50ms\}$$

$$\text{mit } ROR(t) = \frac{E(t)}{E(t-d)} \text{ und } E(t) = \sum_{d=0}^{20ms} x^2(t-d)$$

Als Sprachsignalerkenner kann beispielsweise ein Gaußscher-Klassifikator mit gespeicherten Lautprototypen für die Lautklassen „f", „sch", „ch", „s", „z", „k" und „t" verwendet werden. Der Klassifikator arbeitet vorzugsweise in drei Stufen. In einer ersten Stufe wird zunächst der Eingangswert je Sprachsignalmerkmal mit einem gültigen Bereich gefenstert. Bei den nach dieser Filterung noch verbleibenden Sprachsignalmerkmalen kann mit einem normalverteilungsbasierten Distanzmaß $d_k(x)$ die Entfernung zu den normalverteilten Lautprototypen berechnet werden. Der Lautprototyp, der dem Eingangsvektor, bestehend aus Komponenten unterschiedlicher Sprachsignalmerkmale, am nächsten liegt, wird ausgewählt und dem schwach artikulierten Laut zugeordnet. Zuletzt kann die durchlaufene Entscheidungszeitreihe noch mit einem stochastischen Filter geglättet werden.

Für die Gesamtheit der für eine Lautklasse verwendeten Sprachsignalmerkmale wird zur Fensterung der Eingangswerte je Sprachsignalmerkmal jeweils ein gültiger Bereich zulässiger Werte vorgegeben. Damit eine Lautklasse detektiert werden kann, müssen alle Sprachsignalmerkmale eines Sprachsignals in dem für sie gültigen Bereich liegen. Hierdurch ist eine vorteilhafte unsymmetrische Beschneidung der Füße der Sprachsignalmerkmals-Verteilungen möglich. Des

Weiteren lassen sich dadurch die festen alles-oder-nichts-Bedingungen „muss ROR und Pause haben" für „t", $NCCF_{max} > 0{,}55$ für „z" sowie $NCCF_{max} < 0{,}45$ für alle stimmlosen Lautklassen auf derselben Verarbeitungsebene direkt vorgegeben.

$$x_i \in R_{ik} \forall i$$

Für die nach dem Bereichstest verbleibenden Sprachsignalmerkmalsvektoren kann die Wahrscheinlichkeit der Lautklassenprototypen nach vorgegebenen Gaußdichten ermittelt werden. Wie allgemein üblich wird die Entscheidung mittels eines Distanzmaßes getroffen. Die Lautklasse mit der geringsten entsprechenden Distanz wird als detektiert angenommen. Ohne Kovarianzen berechnet sich die Distanz eines Lautes k über alle Dimensionen i zu

$$d_k(x) = \sum_i (x_{ki} - \mu_{ki})^2 \sigma_{ki}^{-2} + 2\sum_i \log \sigma_{ki}$$

Nach diesen beiden Stufen wird vorzugsweise noch eine zeitliche Glättung vorgenommen. Die oberen beiden Schritte können kontinuierlich ausgeführt werden, so dass pro Eingangs-Sample eine Entscheidung für eine Lautprototypenklasse vorliegt. Alle zeitlichen Entscheidungen können beispielsweise in einen 20 ms Ringbuffer geschrieben werden, von dem der häufigste Klassenwert im 20 ms Intervall als endgültiges Erkennungsergebnis verwendet wird (MAXWINS-Operator).

Der Klassifikator sollte vor seinem Einsatz mit natürlicher Sprache trainiert werden. Zur Ermittlung valider Parameter hat sich das folgende Verfahren bewährt.

Zur Ermittlung der Bereichswerte wird die Funktion „set ranges by agglomeration" genutzt. Dabei wird eine hinreichend variable Stichprobe verwendet. Beispielsweise können Sprecher verschiedenen Alters und Geschlechts jeweils mindestens fünf möglichst verschiedene Äußerungen pro Phonem einsprechen. Die Bereiche können mit dem „erweiterten median" gewonnen werden, wobei die Werte bei 30 und 70 % der sortierten Folge als

Werte eines Segmentes als Grenzen festlegbar sind. Aus der Vereinigungsmenge der Bereiche aller Trainingsworte wird der gültige Bereich je Sprachsignalmerkmal festgelegt. Drei Grenzwerte können beispielsweise nicht trainiert sondern nach Erfahrungswerten vorgewählt werden: Der Bereich des NCC-Maximums wird für stimmlose Laute durch eine obere Schranke von 0,45 und stimmhafte durch eine untere Schranke von 0,55 fest eingestellt. Der Bereich für Pausenlänge wird für Plosive auf mindestens 30 ms festgelegt.

Mit der Erfindung wird weiter vorgeschlagen, dass die synthetischen Laute erzeugt werden, indem eine Rauschsignalkomponente und eine Sinussignalkomponente für einen synthetischen Laut generiert und miteinander kombiniert werden.

Die mehrdimensionalen Gaußverteilungen können direkt aus dem Trainingsmaterial berechnet werden.

Zur Synthese stimmloser synthetischer Frikative als Ersatz für „s“ und „ch“ kann eine Kombination von bandbegrenztem Rauschen mit limitierter Pegelverteilung und einem frequenzvariablen Sinuston gewählt werden. Zudem können gesteuerte Frequenzverschiebungen des addierten Sinustones eingeführt werden, um die Verschiebungen des spektralen Schwerpunktes des ursprünglichen Lautes perzeptiv übertragbar zu machen. Die Formen der Rauschsignale werden vorzugsweise so gewählt, dass eine maximale Ähnlichkeit zum Originallaut trotz veränderter Frequenzen erreicht wird. Dies kann durch spezielle Synthese-Maßnahmen erreicht werden, welche beim Ersatzlaut Wahrnehmungswerte von Schärfe und Rauheit erzeugen, die denen des Originallautes möglichst äquivalent sind trotz geänderter Klangfarbe.

Das Rauschsignal aller Komponenten kann zunächst durch FIR-Filterung von weißem Rauschen (Zufallszahlengenerator) erzeugt werden. Vorzugsweise wird der merkmalstragende Frequenzbereich des Ersatz-„s“ in der Regel bei 1,6 kHz positioniert. Diese Lage erzeugt gute perzeptive Abstände zum natürlichen „s“ und „ch“. Entsprechend kann schmalbandig zwischen 1,4 und 1,8 kHz gefiltert werden. Um die Amplitudenverteilung für eine angenehmste Wahrnehmung von Schärfe zu formen (wenig lärmhaft) kann das resultierende Signal hart begrenzt,

beispielsweise durch sein Vorzeichen ersetzt, und erneut gefiltert werden. Dieser Vorgang wird vorzugsweise mehrfach wiederholt. Als Ergebnis erhält man eine Verteilung der Amplituden mit starker Asymmetrie, d.h. es gibt nur wenige geringfügige Überschreitungen des Limitierungspegels. Ein solches Signal maximiert die Wahrnehmung von Schärfe. Auch bei hohen Präsentationspegeln wird die Erzeugung eines unangenehmen Lärmcharakters vermieden. Ferner werden die Sinneszellen vor hohen Belastungen durch kurze Spitzenpegel geschützt. Das so erzeugte Rauschen kann als Zeitsignal abgespeichert werden. Um Wiederholungsartefakte zu vermeiden, können aus einem hinreichend großen Puffer (etwa 500 ms) zufällig Segmente zufälliger Länge ausgewählt werden. Diese können mittels einer sinusförmigen Überblendung zu einem längeren pseudozufälligen Rauschsignal verkettet werden. Zusätzlich kann eine zweite breitbandige Rauschkomponente für das „s" erzeugt werden. Diese kann durch FIR-Filterung mit einem Durchlassbereich von 800 Hz - 4 kHz aus dem Spektrum weißen Rauschens ausgeschnitten werden. Die vorbeschriebene Formung der Amplitudenverteilung wird auch hier angewendet. Zur Anpassung an den spektralen Verlauf des Hörschadens kann diese Rauschkomponente mit einer entsprechenden Dachschrägenkorrektur versehen werden. Diese zweite Komponente sollte nicht weggelassen werden, da sie dafür sorgt, dass die Bindung des Ersatz-„s" an Kontextlaute mit Merkmalen in diesem Spektralbereich verbessert wird und „stream-segregation" vermieden wird. Sie kann mit einem um ca. -6 bis -12 dB niedrigeren Pegel zur ersten Rauschkomponente addiert werden. Der exakte Pegelwert sollte an den individuellen Hörschaden angepasst werden.

Zur störungsfreien Generierung eines frequenzmodulierten Sinussignals, sollte diese ohne Phasensprünge erfolgen. Dazu kann das Argument der Sinusfunktion durch Integration von Phasenwerten gewonnen werden. Die momentane Frequenz kann aus der Zählung von Nulldurchgängen gewonnen werden. Dazu können im bandpassgefilterten Originalsignal die Nulldurchgänge beispielsweise in einem Zeitfenster der Länge $l = 20$ ms gezählt und der so bestimmte Frequenzschwerpunkt durch eine lineare Kennlinie mit drei Stützstellen in die Frequenz der Sinuskomponente transformiert werden. Die Stützstellenpositionen für „mittleres s", „hohes s" und „tiefes s" können durch einen einfachen Hörtest individuell festgelegt werden.

$$\sin_s(t) = \sin\left(\sum_0^t \phi_s(t)\right)\phi_s(t) = map_s\left(ZCC_s(t)\right)$$

Für eine maximale Naturähnlichkeit des Ersatz-„ch" kann das „ch" beispielsweise aus zwei spektralkonzentrierten Komponenten zusammengesetzt werden. Beispielsweise kann das Ersatz-„ch" aus einer niederfrequenten Komponente um 400 Hz und einer höherfrequenten Komponente um 2 kHz gebildet werden, welche vorzugsweise in Vorversuchen durch Hörexperimente ermittelt worden sind. Ebenso wie beim „s" können diese Rauschsignalkomponenten jeweils durch unterschiedliche spektrale Filterung aus weißem Rauschen erzeugt werden. Eine Änderung der Amplitudenverteilung kann ähnlich wie beim „s" erfolgen: Das Signal kann hier durch zweifache Begrenzung und Filterung so modifiziert werden, dass einerseits hohe Signalspitzen vermieden werden, andererseits jedoch noch wahrnehmbare Fluktuationen, wie auch beim natürlichen „ch", auftreten können. Das so erzeugbare Rauschen kann wiederum als Zeitsignal abgespeichert werden. Um Wiederholungsartefakte zu vermeiden, können aus einem hinreichend großen Puffer (etwa 500 ms) zufällige Segmente zufälliger Länge (mit Längenvariationen zwischen 10 und 20 ms) ausgewählt werden. Diese Segmente können mittels einer sinusförmigen Überblendung zu einem längeren pseudozufälligen Rauschsignal verkettet werden. Dabei kann wiederum eine Besonderheit eingeführt werden: Vorversuche haben gezeigt, dass das natürliche Wahrnehmungsbild des „ch" von den Fluktuationen der Umhüllenden im Bereich von 5 bis 20 ms beeinflusst wird. Diese Wahrnehmung kann beispielsweise Einzelelement-Rauheit genannt werden. Beim synthetisch erzeugten „ch"-Ersatzlaut kann entsprechend Einzelelement-Rauheit erzeugt werden, indem beispielsweise zufällige kurze (5 bis 10 ms währende) Pausen zwischen den Sinusfenstern der einzelnen vorgenannten Rauschsegmente eingeführt werden. Auf diese Art und Weise werden maximale Ähnlichkeiten vom natürlichen „ch" und synthetischen „ch" erreicht. Es wird erwartet, dass auch vom geschädigten Gehör dieses Merkmal gut auswertbar ist.

Ebenso wie beim „s" kann bei der Generierung der Sinuskomponente des Ersatz-„ch" ein Nulldurchgangszähler verwendet werden. Diesmal für das im Bereich von 5 bis 10 kHz bandpassgefilterte Eingangssignal. Der so erhaltene

Wert, welcher als Schätzwert der mittleren Frequenz zu verstehen ist, kann wieder mit einer linearen Abbildungsfunktion transformiert und dann aufaddiert werden, um als Argument einer Sinusfunktion verwendbar zu sein.

$$\sin_c(t) = \sin\left(\sum_0^t \phi_c(t)\right) \phi_c(t) = map_c\left(ZCC_c(t)\right)$$

Das „t" kann durch Einfügung eines Komplexes, bestehend aus einer abgespeicherten synthetischen Impuls-Plosion und einer zusätzlichen breitbandigen Rauschkomponente, welche dem hochfrequenten Teil des synthetischen „s"-Signals gleicht, erzeugt werden. Die abgespeicherte Plosion kann aus einem bipolaren Dreiecksignal gewonnen werden, das beispielsweise mit einem FIR-Bandfilter mit einem Durchlassbereich zwischen 100 Hz und 800 Hz gefiltert wird und das im Abstand von 10 ms zweimal wiederholt wird. Der Einsatzzeitpunkt kann auf den des maximalen ROR gelegt werden. Zur Realisierung der zusätzlichen Rauschkomponente kann für die Dauer der durch die „t"-Erkennung getriggerten „t"-Fensterzeit (50 ms) ein kontinuierliches, breitbandiges, höhere Frequenzen enthaltendes Rauschsignal (800 Hz bis 4 kHz) zugemischt werden. Die Realisierung erfolgt vorzugsweise im Erkenner, der die Signale für Pause und ROR 50 ms hält. Dadurch wird auch die „t"-Aufbereitung im Normalfall 50 ms aufrechterhalten. Der Prozess wird bevorzugt nur dann abgebrochen, wenn die spektrale Form sich stark ändert, derartig, dass die Bandenergiewerte aus dem für „t" erlaubten Bereich fallen. Um den Plosivcharakter zu erhalten, wird der Soft-Switch des „t" beispielsweise mit 2 statt mit 10 ms Einschaltzeit betrieben.

Ferner wird vorgeschlagen, dass die Amplituden der synthetischen Laute vor dem Einfügen in das Sprachsignal individuell komprimiert werden. Hierdurch können die Pegel der synthetisierten Laute den individuellen Recruitment-Kennlinien des geschädigten Ohres angepasst werden. Dazu kann das Originalsignal bandpassgefiltert, der Moving-Average des Betrages gebildet und die resultierende Originalenergie durch eine Kompressions-Kennlinie entsprechend der neuen spektralen Lage transformiert werden. Dazu kann beispielsweise eine 4-segmentige Kompressions-Kennlinie zur Verfügung gestellt werden:

1. Unter th0 wird keine Kompression angewendet.

2. Unterhalb der Schwelle th1 wird keinen Kompression oder eine Expansion angewendet.

3. Oberhalb von th1 wird eine Kompression von r1 (etwa 2 bis 8) : 1 für den Nutzbereich eingestellt.

4. Oberhalb von th2 wird eine begrenzende Kompressionsrate von r2 (etwa 10 bis „unendlich") : 1 oder eine negative Steigung einstellbar gemacht.

Entsprechend wird der Multiplikationsfaktor m abhängig vom Mittelwert x wie folgt berechnet:

$$m(x) = \begin{cases} 1 & x < th1 \\ \dfrac{(r1-1)th1}{r1x} + \dfrac{1}{r1} & th1 < x < th2 \\ \left(\dfrac{th2-th1}{r1r2} + \dfrac{th1-th2}{r2}\right)\dfrac{1}{x} + \dfrac{1}{r2} & th2 < x \end{cases}$$

Das aus Addition der Rausch- und Sinussignale entstandene Summensignal kann mit dem komprimierten Original-Amplitudensignal multipliziert werden. Das Hinzumischen der Ersatzlaute durch das Erkennersignal kann über einen Soft-Switch gesteuert werden, d.h. bei Erkennung eines zu ersetzenden schwach artikulierten Lautes wird dessen Synthese-Signal mit einer über die Dauer einer Einschaltzeit t_{on} (ca. 10 ms) linear ansteigenden Amplitude hinzugemischt. Analog wird das Signal am Ende der Erkennungszeit über eine Ausschaltzeit t_{off} (ca. 20 bis 50 ms) mit linear auf 0 abfallender Amplitude ausgeblendet. Das Eingangssignal wird dabei vorzugsweise gegenüber dem Synthesesignal um 20 ms verzögert, um die Verzögerung durch Erkennen und Einschalten zu kompensieren.

Die Erfindung betrifft weiterhin ein Sprachsyntheseverfahren, insbesondere zur Erzeugung von synthetischen Lauten in einem Verfahren der vorbeschriebenen

Art, wobei zwei oder mehr Formantwellenformen jeweils durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit einer Umhüllenden-Funktion erzeugt werden, die zwei oder mehr Formantwellenformen addiert werden und die addierten Formantwellenformen gemäß einer Pitchintervalllänge und gemäß suprasegmentaler Verkettungsregeln zu einem suprasegmentalen Sprachsignal verkettet werden.

Übliche Verfahren für die Synthese stimmhaft angeregter Sprachsignale gehen von einem stationären Anregungsquellen-Filter-Modell aus. Formantvokoder, die mit stationären Filteransätzen arbeiten – sei es in Seriellstrukturen mit Transversalfilter per Linear Predictive Coding (LPC) oder in Parallelstrukturen mit anderen Koeffizienten gesteuert – können nur in begrenztem Maße die Natürlichkeit der synthetischen Sprache sicherstellen. Natürlichkeit ergibt sich dort quasi zufällig bei günstiger Wahl der Parameter oder erfordert die Einführung spezieller technischer Randbedingungen, z.B. Multi-Puls LPC. Kausale Regeln, welche Natürlichkeit bei der Erzeugung synthetischer Sprache sicherstellen, werden nicht identifiziert.

Die Ursache für starke Unnatürlichkeit bei der Synthese von Sprachsignalen liegen im Auftreten von mit natürlicher Sprache konkurrierenden Perzepten:

1. Wahrnehmung von unnatürlicher Tonalität mit Tonhöhen in den Bereichen der Formantfrequenzen; sich daraus entwickelndes „Zwitschern“ bei schnellen suprasegmentalen Änderungen der Formantfrequenzen;

2. Wahrnehmung von Tonalität mit Tonhöhen im Frequenzbereich der stimmlichen Anregung;

3. Wahrnehmung von Intonations-Betonungs/Rhythmusstörungen bei suprasegmentaler Verkettung von Anregungsperioden-Sequenzen durch inadäquat kontrollierte Modulation der Interferenzenergie bei Variation der Stimmtonhöhe;

4. Wahrnehmung von Identitätsänderungen bei Vokalen durch suprasegmental variierende Stimmtonhöhe bei gleichzeitig konstant gehaltenen Formantfrequenzen.

Bei Reizung des Ohres mit einem Sinussignal wird ein reiner Ton wahrgenommen. Die Qualität dieser Empfindung bezeichnet man mit Tonalität. Sprache enthält keine Tonalität und darf bei synthetischer Herstellung auch keine solche enthalten. Tonwahrnehmungen innerhalb von Sequenzen synthetischer Sprache sind Störungen. Frequenzänderungen von Komplexen, welche Tonalitätsstörungen beinhalten, können besonders störendes „Zwitschern" erzeugen.

Zur Vermeidung dieser „Unnatürlichkeits-Störungen" sind Syntheseregeln einzuführen, die auf Vorgänge bei der natürlichen Vokalartikulation zurückzuführen sind. Sie stellen modelltechnisch gesehen Verfeinerungen der bekannten Quelle-Filter-Synthesemodelle dar. Es liegt eine ungeeignete Vereinfachung bei den klassischen Methoden und Modellen darin, dass das oder die Artikulations-Hohlraumfilter im Verlauf einer stimmhaften Periode nicht nur als fast-stationär, sondern auch als verlustfrei oder zumindest als mit nur konstanten Verlusten behaftet angesehen werden kann, so dass eine lineare Differenzialgleichung mit periodenzeit-invarianten, festen Resonanzfrequenzen und festen Abklingzeiten fälschlicherweise zur Beschreibung verwendet wird. Auch Modelle für das menschliche Hören, deren zentrale Hypothese darin besteht, dass lediglich die Schallleistungsverteilung auf der Basilarmembran die Basis für eine kognitive Signalauswertung ist, leisten der Anschauung Vorschub, dass einfache, stationäre Spracherzeugungsmodelle zur Beschreibung ausreichend seien. Es wird nicht berücksichtigt, dass auch Signalfeinstrukturen in der Pitchperiode differenzierte Wirkungen hervorrufen können.

Die Erfindung schlägt ein Syntheseverfahren vor, bei dem die Quellensignale bei der Erzeugung der Formantwellenformen frequenzmoduliert werden.

Erfindungsgemäß wird somit Tonalität bei repetitiven Wellenformen, die primär aus Sinussignal-Paketen bestehen, durch Frequenzmodulation weitestgehend beseitigt. Mit Hilfe der Frequenzmodulation werden die bei der jeweiligen

Formantfrequenz oszillierenden Quellensignale nach einer vorgegebenen Funktion gewobbelt. Hierdurch liegt eine Frequenzmodulation der ursprünglich frequenzkonstanten Wellenform vor. Die hierdurch variierende Frequenz des Quellensignals verhindert, dass die Basilarmembran im Zeitmittel nur eine schmale Verteilung von Zeitintervallen im akustischen Nerv produziert. Die Verteilung wird durch die Frequenzmodulation verbreitert. Die Frequenzlage des cortikal extrahierten Maximums der Verteilung wird (kontrollierbar) undefinierter. Durch eine geeignete Parameterwahl ist zu erreichen, dass die cortikale Zuordnung der kognitiven Wahrnehmungsklasse „einzelner reiner Ton" nicht mehr möglich ist. Die Wahrnehmung von Tonalität kann weitestgehend geschwächt werden, und es kann insbesondere erreicht werden, dass die Existenz von Tonalitätswahrnehmungen gänzlich verschwindet.

Nach einer vorteilhaften Ausgestaltung der Erfindung ist die Frequenzmodulation der Quellensignale zyklostationär. Diese Art der Frequenzmodulation ist praktisch besonders leicht realisierbar und erzeugt die gewünschte Natürlichkeit der synthetisierten Sprache.

Tonalität kann auch bei intermittierend angebotenen Sinuspaketen quasi hinter der Repetitionstonhöhe herausgehört werden. Das gilt besonders auch bei den stimmperiodisch repetierten Sinus-Bursts von Formantwellenformen. Das Perzept Tonalität kann somit über kurze Pausen integriert werden. Bei verkürzter Tastzeit schwächt sich dieses Perzept ab und verschwindet quasi in der Stärke der Perioditäts-Pitch-Wahrnehmung.

Nach einer weiteren vorteilhaften Ausgestaltung der Erfindung ist vorgesehen, dass bei der Überlagerung und Verkettung der Formantwellenformen keine störenden Interferenz-bedingten Umhüllendenschwankungen auftreten. Das bedeutet, dass die Überlappung aufeinander folgender Pitchwellenformen kontrolliert werden muss, was durch Pitch-abhängige Adaption der Wellenpaketlänge, durch Modifikation der Längen H und A (vgl. Fig. 1) erreicht wird. Vorzugsweise erfolgt die Verkettung der frequenzmodulierten Wellenpakete derart durch pitchadaptive Umhüllendenformung, dass keine wahrnehmbaren Störungen durch Modulation im Überlagerungsbereich der Wellenzüge auftreten.

Gemäß einer weiteren vorteilhaften Ausgestaltung der Erfindung hängt der Modulationshub der Frequenzmodulation bei der Erzeugung der Formantwellenformen von der jeweiligen mittleren Formantfrequenz ab. Da das erfindungsgemäß frequenzgewobbelte Sinuspaket einen optimal klassifizierbaren Vokalformanten darstellen soll, kann die Frequenz eines Quellensignals nicht beliebig weit von der ursprünglichen Sinusfrequenz ausgelenkt werden. Es darf nicht passieren, dass der kognitive Bereich des „guten Vokalprototypen" verlassen wird. Dies kann durch entsprechende Vorgaben von Bereichsfunktionen erreicht werden. Bei natürlicher Artikulation enthält eine Formantfrequenz innerhalb einer Periode starke Mikro-Schwankungen. Dies mag der Grund dafür sein, dass bei natürlicher Artikulation Tonalität niemals ein Problem ist. Die Ausdehnung der Existenzbereiche der Vokale, soweit diese von zwei Formanten ohne Variation der Frequenz der Quellensignale aufgespannt werden, kann durch psychophysische Versuche zuvor ermittelt werden. Diese Ausdehnung der jeweiligen Existenzbereiche beider Formanten hängt im Wesentlichen von der mittleren Lage der Formanten auf der Basilarmembran ab. Bei der Synthese von 2-formantigen Vokalen können für die beiden oszillierenden Quellensignale beispielsweise die folgenden zwei Bereichsfunktionen vorgegeben werden: Eine für erste Formanten im Bereich von bis zu 1000 Hz und eine für zweite Formanten im Bereich von 500 Hz bis 4 kHz. Der erlaubte Wobbelbereich nimmt stark mit steigender Frequenz zu.

Zweckmäßigerweise beträgt der Modulationshub der Frequenzmodulation bis zu 20%, vorzugsweise bis zu 10% der jeweiligen mittleren Formantfrequenz. Bevorzugt ist der Modulationshub der Frequenzmodulation bei der Synthese weiblicher Sprache kleiner als bei der Synthese männlicher Sprache. Die typische Abweichung für männliche Sprecher liegt beispielsweise für breite u-Formanten unter 200 Hz bei konstant 10%, fällt dann (prozentual) linear bis 1 kHz und steigt bis 4 kHz wieder leicht an. Bei hohem Pitch weiblicher Sprecher kann weniger Frequenzmodulation verwendet werden. Es wird beispielsweise die bei Männern gewählte prozentuale Abweichung halbiert.

Eine weitere vorteilhafte Ausgestaltung der Erfindung sieht vor, dass bei der Überlagerung und Verkettung der addierten Formantwellenformen die

Pitchintervalllänge variiert wird. Bei suprasegmentalen Verkettungen von stimmhaften Segmenten wird bevorzugt eine randomisierte Variation der Pitchintervalllänge eingeführt, wobei die maximal vorkommende Abweichung vorab festlegbar ist. Diese Ausgestaltung dient der Vermeidung des Auftretens von Tonalität bei äquivalenter Synthese stimmhafter Pitch-Anregungs-Intervalle. Eine exakt wiederholte Pitchwellenform erzeugt bei neuronaler Auswertung der Repetitionsintervalle eine sehr schmale und energiereiche Häufigkeitsverteilung der Pitchintervall-zugeordneten Impulsspikes im akustischen Nerv; denkbar als Kreuzkorrelation. Die resultierende Stärke der Wiederhol-Häufigkeitsverteilung bei einer Synthese mit exakt gleichen Pitch-Wiederholintervallen ist sprachuntypisch, da eine < 1%-ige Genauigkeit der Repetition über mehrere Perioden beim natürlichen Sprechen nicht vorkommt. Insgesamt soll somit eine Langzeitkonstanz bei suprasegmentalen Pitchkonturen vermieden werden. Nach vorteilhafter Ausgestaltung der Erfindung wird die Pitchintervalllänge so variiert, dass ihr momentaner Wert mit stochastischen Schwankungen versehen wird, die maximal 1% bis 2% bei der Synthese männlicher Sprecher, aber nur <0,5% bei der Synthese weiblicher Sprecher betragen.

Eine weitere vorteilhafte Ausgestaltung sieht zusätzlich eine Regel vor, nach der eine absolute Konstanz des stilisiert synthetisierten Pitchverlaufs (ohne die vorgenannten stochastischen Schwankungen) über ein typisches Silbenintervall (ca. 200 ms) untersagt wird; die Abweichung von einem Horizontalverlauf muss hier >3% betragen.

Es wird weiter vorgeschlagen, dass die Umhüllenden-Funktionen aus drei zeitlich aufeinander folgenden Segmenten, nämlich einem Einschwingsegment, in dem die Amplitude des Quellensignals von Null ausgehend ansteigt, einem Haltesegment, in dem die Amplitude des Quellensignals konstant ist, und einem Ausschwingsegment, in dem die Amplitude des Quellensignals wieder auf Null abfällt, besteht. Die Fensterung des Quellensignals durch das Einschwingsegment wird vorzugsweise als Funktion der Formantfrequenz gewählt. Die zugrunde liegende Modellvorstellung ist, dass bei natürlicher Artikulation das Einschwingsegment durch das abrupte Schließen der Glottis ausgelöst wird. Die Umhüllendenanstiegssteilheit ist durch die „Filtergüte“ des Hohlraums bei geschossener Glottis gegeben. Die Güte variiert mit der

Formantresonanzfrequenz. Vorzugsweise ist auch die zeitliche Länge des Haltesegmentes von der Frequenz abhängig. Diese Phase wird approximativ als weitgehend konstanter Zustand der Hohlraumschwingung bei geschlossener Glottis aufgefasst. Auch das Ausschwingsegment wird analog zum Einschwingvorgang mit einem Fenster versehen, dessen Länge vorzugsweise von der Frequenz des Quellensignals abhängig gemacht wird. Während des Ausschwingsegmentes ändert sich der Zustand des Systems, so dass mit unterschiedlichen, variierenden Verlusten gerechnet werden muss, die ihrerseits das Ausschwingsegment beeinflussen können. Diese Systemannahme wird verwendet, um später das Ausschwingsegment als Funktion der Pitchfrequenz naturnah zu variieren. Um unerwünschte Tonalität zu verhindern und maximale Natürlichkeit zu erreichen, sollten die Segmente der Umhüllenden-Funktion als Funktion der Frequenzen der Quellensignale beispielsweise wie folgt verändert werden: Für das Haltesegment werden lineare Segmentfunktionen in drei Trägerfrequenzbereichen verwendet. Das Ausschwingsegment wird als prozentualer Anteil der Pitchperiode festgelegt. Der prozentuale Anteil ist eine Funktion der Frequenz des Quellensignals, die vorzugsweise unter 800 Hz konstant gewählt wird und darüber linear bis 4 kHz abfällt.

Gemäß einer weiteren vorteilhaften Ausgestaltung der Erfindung hängen die Dauer des Einschwingsegments, des Haltesegments und/oder des Ausschwingsegments von der Pitchintervalllänge ab. Zweckmäßigerweise wird bei sinkender Pitchintervalllänge zunächst die Dauer des Ausschwingsegmentes bis auf einen Minimalwert verkürzt und dann die Dauer des Haltesegments verkürzt, so dass Interferenzen von Formantwellenformen aufeinander folgender Pitchintervalle vermieden werden. Durch eine kaskadierende Verkürzungsstrategie wird erreicht, dass zunächst keine ungewollte Bandbreitenerhöhung stattfindet. Nach Verkürzung des Ausschwingsegmentes wird bei weiterem Anstieg der Anregungsfrequenz das Haltesegment der Formantwellenform verkürzt, im Grenzfall verschwindet das Haltesegment vollständig. Die Dauer des Einschwingsegmentes entspricht vorzugsweise einer ganzzahligen Anzahl von Nulldurchgängen der Oszillationen des Quellensignals. Die Anzahl der Nulldurchgänge ist als Funktion der Formant-Mittelfrequenz festgelegt. Sie steigt vorzugsweise bis 1 kHz an, um ein realitätsnahes Einschwingverhalten niedriger Formanten zu erhalten. Von 1 bis 2,6 kHz steigt sie vorzugsweise flacher weiter

bis 3 kHz an und fällt danach mit hoher Steigung wieder ab. Es wird so verhindert, dass Perioden mit resultierendem unnatürlich überdeutlichem Perzept des zweiten Formanten auftreten – wenn denn ein naturnahes Perzept und kein überdeutliches gewünscht wird. Wenn allerdings letzteres doch gewünscht wird, um die Verständlichkeit in Gegenwart von Lärm zu erhöhen, kann auch eine „Überdeutlich"-Einstellung gewählt werden.

Nach klassischen Modellen gibt es keine in der Pitchperiode variierende Verlust- oder Absorptionsfunktionen. Dann sind die unabhängig gedachten Formantwellenformen linear zu superponieren. In diesen Fällen würden bei großer Überlappung zwischen den Perioden bei fester Artikulationskonfiguration Anregungsfrequenz-abhängige Interferenzmodulationen als Funktion des Anregungsfrequenzverlaufes resultieren. Das ist bei realer Sprache nicht festzustellen. Zur Unterdrückung solcher Interferenzmodulationen von Formantwellenformen aufeinander folgender Pitchintervalle wird daher vorzugsweise eine Wellen-Dämpfung durchgeführt, welche beispielsweise als Funktion der Glottis-Öffnung das Halte- und das Ausschwingsegment beeinflusst. Von Vorteil ist, wenn das Ausschwingsegment der Umhüllenden-Funktion so ausgebildet ist, dass die Amplitude zum Ende des Pitchintervalls auf höchstens 35%, vorzugsweise auf höchsten 25% der während des Halte-segments konstanten Amplitude abgefallen ist. Der Wert der Endamplitude wird vorzugsweise Pitch-adaptiv festgesetzt.

Es wird weiter als vorteilhaft erachtet, wenn das Sprachsignal eine Hochpassfilterung erfährt. Vorzugsweise erfolgt eine Hochpassfilterung mittels eines IIR-Filters mit einer Cutoff-Frequenz von 100 Hz. Hierdurch können unerwünschte niederfrequente Signalanteile eliminiert werden, die durch Superposition von Wellenformen bei variabler Pitchintervalllänge entstehen.

Sofern gemäß der Erfindung 2-Formant-Vokale synthetisiert werden sollen, muss noch der relative Amplitudenunterschied der beiden festgelegt werden. Das Pegelverhältnis wird als zweidimensionale Funktion abhängig von den Frequenzen der ersten und der zweiten Formanten (F1 bzw. F2) definiert. Die unten dargestellte Tabelle zeigt Werte für typische Vokallagen. Auf dieser Basis können für Zwischenlagen Verhältniswerte aus den tabellierten Stützstellen

interpoliert werden. Dazu wird einmalig die Triangulation der F1/F2-Stützpunkte berechnet und dann jeder benötigte Wert als Punkt auf einer Ecke-Seiten-Strecke des umgebenden Dreiecks berechnet. Die Werte werden dabei durch Abgleich des sich ergebenden Synthese-Spektrums mit dem Spektrum von natürlichen Lauten nach der Vorgabe aller anderen Parameter ermittelt. Der F1/F2-Wert für „i" (200/2800) wurde mit -12 dB eher hoch geschätzt, um die Synthese nicht unnötig schlecht verständlich zu gestalten. Die folgende Tabelle gibt passende Vorgaben an:

Vokal	F1 [Hz]	F2 [Hz]	E2/E1 [dB]
U	250	700	-12.00
O	380	600	-6.00
A	800	1250	-6.00
I	200	2800	-12.00
Ä	578	2019	-4.00
E	330	2300	-9.00
Ü	210	1700	-15.00
Ö	365	1400	-10.00
@	500	1200	-9.00
0	550	600	-9.00

Klassische Formantvokoderansätze haben als Erkenntnis die Beschreibung populär gemacht, dass für die Kognition von stimmhaft angeregten Lauten, besonders Vokalen, die Dimensionalität des Klang-Klassifizierraums eine Größenordnung von nicht mehr als 2 besitzt. Wobei die kognitiv klangwirksamen höheren Formantlagen zum Teil durch perzeptive Fusion von mehreren Artikulations-Resonanz-Formanten entstehen. Um eine für Sprachausgabe akzeptable Natürlichkeit zu erreichen, müssen nicht nur die Maxima der Synthesefilter exakt definiert sein, sondern auch die spektralen Abklingfunktionen. Unklar ist, welches kognitive Teil-Perzept durch unterschiedliche Resonanz-„Füße" beeinflusst wird. Untersuchungen haben gezeigt, dass durch die Energieverteilung in den spektralen Tälern des Sprachsignals wesentlich die Wahrnehmungsstärke der Anregungs-Repetition, also der Stimmtonhöhe (pitch) bestimmt wird. Z.B. trägt bei „ä" die Spektrumsgegend unterhalb des zweiten Formanten besonders stark zum Gesamtperzept des „ä" bei. Auch beim „a" ist substantielle Tal-Energie erforderlich, um akzeptable Natürlichkeit zu bekommen. Um nicht hohl zu klingen, benötigt das „a" substantielle Energie unterhalb des ersten Formanten. Um also die Wahrnehmungen „Hohlheit" und „fehlende Repetitions-Stärke" zu vermeiden, wird gemäß der Erfindung vorzugsweise, vor allem für bessere Natürlichkeit bei den Vokalen „ä" und „a", eine dritte Synthesekomponente eingeführt: Außer dem ersten und dem zweiten Formanten wird in begrenzten Bereichen der Vokal-Ebene, die aus eben diesen aufgespannt wird, zusätzliche Signalintensität mit lageabhängigem Pegel und Frequenzgehalt hinzugefügt. Diese dritte Komponente kann ein gefilterter Kunstglottisimpuls sein, der aus einem Basisimpuls erzeugt wird. Er enthält somit spektral bestimmte repetitive Tal-Energien, die unterschiedliche resonanzfreie, bandbegrenzte Pitch-Linienspektren in definierten unterschiedlichen Bereichen der F1/F2-Ebene (Vokalkarte) aufweisen. Dazu kann ein Anregungsimpuls mit einem FIR-Filter gefiltert werden. Um ein vorgegebenes Zielspektrum zu erreichen, wird dieses FIR-Filter nach der Vorgabe H(f) abzüglich des Eigenspektrums des Impulses $H_I(f)$ gewählt, so dass das Spektrum des gefilterten Impulses möglichst genau der Vorgabe entspricht. Das vorgegebene Spektrum kann beispielsweise im Falle des „a" ein Tiefpass mit einer Eckfrequenz bei 400 Hz sein. Als Impuls kann eine einzelne Sinus-Schwingung mit T=2ms gewählt werden.

Das erfindungsgemäße Verfahren ermöglicht die Synthese „superdeutlicher" Vokale, was z.B. vorteilhaft zur Erzeugung von Testsignalen für die Anpassung von Hörgeräten ist. Weiterhin sind solche Vokale für Personen mit Hördefiziten besser verständlich. Durch die beschriebene erfindungsgemäße Vorgehensweise wird es nämlich möglich, das Signal-Rausch-Verhältnis der merkmalstragenden Komponenten des Sprachsignals und damit die neuronale Korrelationsstärke pro Vokalklasse zu maximieren. Die spektrale Talabsenkung und damit der spektrale Modulationsgrad werden so weit getrieben, wie dies die Natürlichkeits-Randbedingungen zulassen. Die Vokal-definierenden Formanten werden mit Formant-Fenstern mit maximaler temporaler Länge bei adäquater Frequenzmodulation erzeugt. Durch die Zeitverlängerung bei höheren Formanten relativ zu natürlichen Fensterlängen und durch die Konzentration der merkmalstragenden Energien mit maximaler Ausdehnung in den perzeptiv effektiven spektralen Merkmals-Detektions-Bereichen lassen sich superdeutliche bzw. super(rausch)störungsresistente Vokalprototypen erzeugen, wobei sich ein besonderer Vorteil bei Sprachausgabe mit solchermaßen erzeugten Vokalen in gestörten Umgebungen ergibt.

Mit der Erfindung wird ferner vorgeschlagen, dass bei Änderung der Pitchintervalllänge die (mittleren) Formantfrequenzen pitch-abhängig variiert werden, und zwar in der Weise, dass die Formantfrequenzen bei Verkürzung der Pitchintervalllänge erhöht werden. Hierdurch können natürlich intonierte, rhythmusbehaftete, suprasegmentale Sequenzen erzeugt werden, in denen eine naturnahe perzeptive Vokalstabilität gewährleistet ist. Für eine als natürlich empfundene Synthese werden allein die messbaren erforderlichen Formantänderungen als Funktion von Pitchänderungen benötigt, die für den optimalen Identitätserhalt des Vokalbildes erforderlich sind. Wenn bei deutlichen Variationen, besonders bei Hebungen, der Stimmtonhöhe die mittlere Formantfrequenzlage konstant beibehalten wird, kann dies nicht nur zu einem Eindruck von Unnatürlichkeit führen, die Wahrnehmung kann bei erheblichen Verschiebungen der mittleren Pitchlage eine Klassengrenze überspringen, so dass der Vokal perzeptiv in eine andere Klasse mutieren kann (Mann-Frau-Kind-Sopran). Im Bereich der umgangssprachlich auftretenden Pitch-Intonationsvariationen, einschließlich der Mann-Frau-Unterschiede, findet man also, dass zur Verhinderung dieser Störungen der Vokal-Wahrnehmungskonstanz die

Formantlagen im suprasegmentalen Zeitmaßstab nach eindeutigen Funktionen verändert werden müssen. Der perzeptive und kognitive Mechanismus, der der festgestellten Vokalkonstanz zugrunde liegt, ist bislang nicht vollständig aufgeklärt. Es gibt nach derzeitigen Evaluierungen mehrere Mechanismen und Transformationen, deren Wirksamkeit vermutet werden muss. Bei Synthese isolierter Vokale mit einfacher, fast konstant gehaltener, einsegmentiger Pitchkontur kann die Formantfrequenz in gleicher Art und Weise wie bei komplexen suprasegmentalen Pitchkonturen mitvariiert werden. Bei Einführung von intonationsbedingten komplizierteren Pitchkonturen ist es wesentlich, dass die Formantfrequenzen in die gleiche Richtung variiert werden, die von der Pitchänderung vorgegeben wird. Dazu kann beispielsweise eine Mitkopplung von 1 bis 5% Formantfrequenzänderung bei 10% Pitchänderung im Suprasegment verwendet werden. Bei Plosiven ist die spektrale Bewegung der Formanten hin zum zentralen Plosiv-Frequenz-Schwerpunkt bzw. Maximum quasi als Nachahmung der Koartikulation natürlichkeits- und klarheitsfördernd und kann mit diesem Verfahren über die Formantkorrektur-Funktion, die bislang für die Pitchanpassung zuständig ist, ebenfalls eingeführt werden.

Gemäß einem weiteren Aspekt betrifft die Erfindung ein Verfahren zur Kontrolle der Anpassung eines Hörgerätes, insbesondere eines Hörgerätes, dessen Funktion auf dem vorbeschriebenen Verfahren basiert, wobei das Hörgerät eine Filterbank zur spektral selektiven Verstärkung und Dynamikkompression von Audiosignalen aufweist, an ein Hördefizit eines Hörgeräteträgers, wobei mittels einer Signalquelle ein Testsignal erzeugt wird und die Wahrnehmung des Testsignals durch den Hörgeräteträger bewertet wird.

Moderne Hörgeräte erlauben es prinzipiell, Hördefizite individuell gut zu kompensieren. Um eine individuell optimale Anpassung des Hörgerätes an das Hördefizit des Hörgeräteträgers zu erreichen, sind jedoch eine Vielzahl von Parametern des Hörgerätes einzustellen und präzise Kontrollen durchzuführen. Es sind dies die Verstärkungs- und Kompressionsparameter der für die verschiedenen Spektralbereiche zuständigen Filter der Filterbank des Hörgerätes. Häufig ist die nötige Zeit oder sind die notwendigen Mittel zur hinreichend Kontrolle und Anpassung des Hörgerätes nicht verfügbar. Es ist festzustellen, dass die Qualität der Anpassmethoden in der jüngeren

Vergangenheit nicht Schritt halten konnte mit der Entwicklung der Hörgerätetechnik und generell der technischen Verarbeitungsmöglichkeiten von Audiosignalen, insbesondere von Sprachsignalen. Daher findet man oft suboptimale Anpassungsresultate.

Eine suboptimale Anpassung eines Hörgerätes an das individuelle Hördefizit des Hörgeräteträgers hat nicht hinnehmbare Auswirkungen auf die Kommunikationsfähigkeit des Hörgeräteträgers, vor allem, wenn hochgradige sensorische Schäden mit stark eingeengten Dynamikbereichen vorliegen. Bei solchen Schäden dürfen die Anpassungskriterien nicht auf einen allgemeinen Ausgleich von Verstärkungsfaktoren im Spektrum ausgerichtet sein. Stattdessen muss die Anpassung gezielt auf die Wiederherstellung der Sprachkommunikationsfähigkeit in den wichtigsten Gesprächssituationen (ggf. mit der jeweiligen Störungsumgebung) gerichtet sein.

Heutzutage ist es üblich, zur Kontrolle der Anpassung eines Hörgerätes den Hörgeräteträger mit Testsignalen zu beaufschlagen, die mittels einer geeigneten Signalquelle erzeugt werden. Die Wahrnehmung des jeweiligen Testsignals wird durch den Hörgeräteträger bewertet. Das Bewertungsergebnis lässt auf die Güte der Anpassung des Hörgerätes zurück schließen. Dementsprechend kann dann die Einstellung des Hörgerätes geändert werden.

Da es bei der Anpassung des Hörgerätes vor allem um die Sicherstellung der auditiven Übertragung von Sprache geht, werden zur Kontrolle der Anpassung häufig so genannte Sprachverständlichkeitstests angewendet, z. B. auf der Basis von sprachrepräsentativen Wortlisten. Es zeigt sich jedoch, dass die Ergebnisse solcher Sprachverständlichkeitstests keine ausreichend differenzierten Rückschlüsse auf die vorzunehmende Einstellung der Verstärkungs- und Kompressionsparameter ziehen lassen.

Durch die Erfindung soll ein Verfahren zur Kontrolle der Anpassung eines Hörgerätes bereitgestellt werden, welches es ermöglicht, entsprechend den Verarbeitungsmöglichkeiten moderner digitaler Hörgeräte mit Filterbänken und pro spektral selektivem Filter einstellbaren Verstärkungsfaktoren und

Kompressionskennlinien, durch entsprechende Einstellung des Hörgerätes, eine optimale Sprachkommunikationsfähigkeit herzustellen.

Diese Aufgabe löst die Erfindung dadurch, dass das Testsignal wenigstens ein natürliches oder naturähnliches Sprachelement umfasst, das in der Weise spektral gefiltert oder ausgewählt ist, dass das Spektrum des Testsignals zu dem Spektralbereich wenigstens eines Filters der Filterbank des Hörgeräts korrespondiert.

Bei zunehmenden Gehörschädigungen nimmt die Redundanz in der Menge der verwertbaren Sprachsignalmerkmale stark ab. Oft sind Total-Hör-Ausfälle einer zunehmenden Anzahl von Sprachbestandteilen zu verzeichnen. Daher ist es für die zu fordernde optimale Ausnutzung des Gehörs von entscheidender Wichtigkeit, dass relevante Sprachbestandteile, die für die Sprachverständlichkeit wichtig sind, in einem nutzbaren Sprecher-Abstandsbereich gut hörbar gemacht werden. Nichthörbarkeit von Sprachelementen oder merkmalstragenden Bestandteilen im Resthörbereich oder die permanente Existenz von zu lauten Schallen sind schwere Anpassungsmängel. Diese Mängel müssen unbedingt ausgeschlossen werden.

Im Prinzip müssten sämtliche Phoneme und ggf. deren spektro-temporalen Bestandteile und die wichtigsten Logatom-Kombinationen und Silben, insofern sie phonetisch-linguistische Ganzheiten bilden, der jeweiligen Sprache überprüft werden. Grundlage kann z. B. die IPA-Liste (International Phonetic Alphabet) zusammen mit Logatom-Listen der CCITT sein. Ein solcher Aufwand wäre allerdings für eine praktikable Kontrolle der Anpassung eines Hörgerätes viel zu hoch. Gemäß der Erfindung werden stellvertretend Testsignale verwendet, die natürliche oder naturähnliche Sprachelemente sind.

Die Überprüfung und Einstellung des Höregerätes kann gemäß der Erfindung mit gefiltertem (natürlichen) Sprachmaterial erfolgen. Stattdessen können auch eindeutig definierte, sprecherunabhängige Testsignale verwendet werden. Diese müssen allerdings möglichst naturähnlich erzeugt werden. Hierfür eignet sich das weiter oben beschriebene erfindungsgemäße Sprachsyntheseverfahren besonders gut.

Vorzugsweise werden zur Überprüfung und Einstellung des Höregerätes solche Sprachelemente verwendet, die als Problemfälle ohnehin bekannt sind. Es sind dies artikulativ zu schwache Elemente, z. B. /ch, s, f, sch/ oder artikulativ zu starke Elemente, z. B. /a, ä/, oder auch sehr kurze und schwache Elemente, z. B. /p, t, k, b, d, g/, oder Sprachelemente, deren merkmalstragende Energie außerhalb des Resthörbereichs liegt z. B. /s, z, c, h, t/. Die Auswahl der vokalischen Sprachelemente als Testsignale oder die spektrale Filterung erfolgt gemäß der Erfindung so, dass die Testsignale kritische Bereiche der Übertragung im Spektrum abdecken, so dass aus der Bewertung der Wahrnehmung der entsprechenden Testsignale durch den Hörgeräteträger gezielt Rückschlüsse auf suboptimal eingestellte Parameter der Filterbank des Hörgerätes gezogen werden können. Dies ist gemäß der Erfindung möglich, weil das Spektrum der Testsignale zu dem Spektralbereich wenigstens eines Filters der Filterbank des Hörgeräts korrespondiert. Damit ist gemeint, dass die Testsignale eine gewisse spektrale Konzentriertheit aufweisen, so dass gezielt darauf geschlossen werden kann, welche Geräteparameter oder welche Parametergruppen des jeweiligen Hörgerätes nicht optimal eingestellt sind. Die Erfindung verlangt nicht zwingend, dass die Testsignale hinsichtlich ihres Spektrums 1:1 auf die spektrale Konfiguration der Filterbank des Hörgerätes abgestimmt sind. Wichtig ist, dass die Testsignale auch nach Filterung der zu Grunde liegenden Sprachelemente noch als Sprache wahrgenommen werden.

Es ist hervorzuheben, dass die herkömmliche regelbasierte Wahl der Hörgeräteparameter oft suboptimal ist. Der Grund dafür ist die Tatsache, dass bekanntermaßen keine festen verallgemeinerungsfähigen Beziehungen zwischen Reintonaudiogramm und komplexer neurophysiologischer Sprachmerkmalsübertragung bestehen. Die heutzutage regelbasierte Ableitung der Geräteparameter aus Audiogrammschwellen erlaubt nur eine rudimentäre Individualanpassung. Die danach häufig vorgenommene Nachjustierung der Geräteparameter „per Hand“ erreichen oft nicht die anzustrebende Wiederherstellung der Sprachkommunikationsfähigkeit. Die Lösung dieser Problematik liegt gemäß der Erfindung darin, als Testsignale gezielt natürliche oder naturähnliche Sprachelemente zu verwenden, wobei die Testsignale eine spektrale Selektivität aufweisen, in der Weise, dass die Bewertung der

Wahrnehmung der Testsignale durch den Hörgeräteträger eindeutige Rückschlüsse auf die nicht optimal eingestellten Geräteparameter ermöglichen.

Bei Anwendung des erfindungsgemäßen Verfahrens sollte geprüft werden, ob die Gesamtheit der in das Restgehör abbildbaren (wesentlichen, repräsentativen) Sprachmerkmale in praxisgerechten Entfernungsbereichen in den verschiedenen Kommunikationssituationen auch wirklich mit angenehmer, verwertbarer Lautheit, d. h. einer solchen Lautheit, die Unterscheidbarkeit und Klarheit erzeugt, zur Verfügung steht. Dies muss in den relevanten Kommunikationssituationen abgeprüft werden. Zum Beispiel muss die Erzeugung der Testsignale und entsprechende Bewertung durch den Hörgeräteträger für Partner-Sprache in üblichen Kommunikationsentfernungen von 0,5 bis 2 m, vorzugsweise 1 m folgen. Die entsprechenden Pegel der Testsignale können durch Signal-Schalldruckpegel-Messungen natürlicher Sprache ermittelt werden. Außerdem muss für die eigene Sprache als Basis einer guten Eigen-Artikulationskontrolle der Abstand zwischen dem Mikrophon des Hörgeräts (z. B. hinter dem Ohr) und dem Mund (ca. 15 cm) simuliert werden. Schließlich können situationsbedingte Testsignale für Sprecher in größeren Entfernungen, z. B. in Vortragssituationen, generiert werden. Der Hörgeräteträger beurteilt die Wahrnehmung der jeweiligen Testsignale und gibt vorzugsweise eine abgestufte Bewertung ab. Dabei gibt er z. B. an, ob das jeweilige Testsignal zu laut, laut, angenehm, leise oder zu leise wahrgenommen wird. Als quantitatives Maß für die individuelle Güte der Anpassung kann dann die Gesamtheit der nutzbaren Entfernungsbereiche und ihre Schnittmenge betrachtet werden, in denen eine kognitiv verwertbare Übertragung der Sprachelemente erzielt werden kann.

Zentrales Merkmal der Erfindung ist somit die Verwendung von Testsignalen, die Sprachelemente sind und gleichzeitig spektral konzentriert sind. Die Testsignale sollten in vom Nutzer testweise kontrollierbaren variablen Entfernungen prüfbar gemacht werden, um eine gute Anpassung in allen relevanten Kommunikationssituationen sicherzustellen. Die spektrale Konzentriertheit der Testsignale ermöglicht eine gezielte Einstellung des Hörgeräts nach Maßgabe der Bewertung der Testsignale durch den Hörgeräteträger. Hierzu muss eine Mehrzahl von Testsignalen erzeugt werden,

die entsprechend die unterschiedlichen Spektralbereiche der Filterbank des Hörgeräts abdecken.

Die Besonderheit der Testsignale gemäß der Erfindung ist, wie bereits erwähnt, dass sie zum einen Sprachelemente sind, d. h. Sprachcharakter besitzen, und gleichzeitig größtmögliche spektrale Konzentriertheit aufweisen. Ein wesentliches Problem bei der Anpassung eines Hörgerätes ist, dass in ein und demselben Spektralbereich große Pegelunterschiede dynamisch korrekt abgebildet werden müssen. Dies gilt insbesondere bei den zweiten Formanten von /i/ und /ä/. Es ist speziell zu prüfen, ob spektral hoch liegende, schwache Merkmale in akzeptablem Abstand hinreichend verarbeitet werden, so dass sie verwertbar laut gehört werden. Es ist andererseits zu prüfen, ob nicht im gleichen Spektralbereich zu hohe Lautheitspegel vorkommen, die schlicht unangenehm laut sind oder die eine unvertretbar starke Maskierung von Nachbarphonemen nach sich ziehen. Hierzu sollte jeweils die weiteste und die kürzeste Entfernung ermittelt werden, die noch gerade nicht zur Störungen der Wahrnehmungen führt. Zweckmäßigerweise sollte das jeweils gleiche Testsignal wiederholt mit unterschiedlicher Lautstärke erzeugt werden, wobei nach Maßgabe der Bewertung durch den Hörgeräteträger die Kennlinien der spektral selektiven Dynamikkompression des Hörgerätes eingestellt werden.

Sinnvollerweise sollte bei dem erfindungsgemäßen Verfahren in einem weiteren Verfahrensschritt ein Testsignal erzeugt werden, dass einem natürlichen Frikativ entspricht. Solche Testsignale sind spektral weit ausgedehnt, mit Merkmalsenergien, die oft weit außerhalb des nutzbaren Resthörbereichs liegen, so dass nur sehr schwache Restenergien in den Resthörbereich fallen. Zu prüfen ist, ob diese kognitiv hinreichend verwertbar gemacht werden. Auch hier sollte jeweils die weiteste und die kürzeste Entfernung ermittelt werden, die noch gerade nicht zu Einbußen der Wahrnehmung führt. Bei der Bewertung durch den Hörgeräteträger sind ggf. spezifische Fragen zu stellen, wie z. B.: Welche dieser Laute sind überhaupt hörbar? Wie ist ihre Unterscheidbarkeit? Häufig setzt die mittels des Hörgerätes maximal mögliche Verstärkung in den oberen Frequenzbändern die Entfernungsgrenze für Frikative auf zu geringe Abstände. Das Rückkopplungspfeifen setzt bei ungünstigen Konstruktionen oder bei Undichtigkeiten der Otoplastik schon bei geringen Verstärkungen ein, die für

eine ausreichende Verstärkung der frikativischen Energien zu gering sind. In diesem Fall muss ein anderes Hörgerät gewählt werden oder die akustische Anpassung und Dichtigkeit der Otoplastik muss verbessert werden. Aus der Bewertung durch den Hörgeräteträger können sich auch Zusatzanforderungen an die Technik des Hörgerätes ergeben, wie beispielsweise eine zusätzliche selektive Sprachmerkmalsverstärkung oder auch eine spektrale Transposition.

Ähnliches gilt für Plosive, die gemäß der Erfindung ebenfalls in einem weiteren Verfahrensschritt abgeprüft werden sollten. Diese Sprachelemente sind Kurzzeitreize mit Impulscharakter mit koartikulativ verteilten Merkmalen, die zum Teil sehr geringe Pegel aufweisen, d. h. sie gehen oft in den Umweltgeräuschen unter. Die Spektren der Plosvie sind ausgedehnt. Daraus resultiert wiederum, dass bei Hochtonverlusten große Teile aus dem Restgehör herausfallen, also nicht verwertbar sind. Die Bewertungsfragen sind also ähnlich wie bei Frikativen. Aus der Bewertung kann sich ergeben, dass eine höhere Grundverstärkung in hohen Frequenzbändern (>2000 Hz) erforderlich ist. Gegebenenfalls muss eine Hörgerätetechnik mit spektraler Transposition verwendet werden. Die Verwendung von natürlichen Plosiven als Testsignale ist außerdem zweckmäßig, da ausgenutzt werden kann, dass Plosive (durch kooperative Gesprächspartner) überdeutlich gesprochen werden können. Falls sich bei der erfindungsgemäßen Überprüfung der Anpassung des Hörgerätes ergibt, dass die Plosiv-Wahrnehmung schlecht ist, kann sich daraus für den Hörgeräteträger die Erkenntnis ergeben, dessen Gesprächspartner zu veranlassen, entsprechend überdeutlich zu artikulieren. Hieraus ergibt sich nachweislich ein hoher Gewinn für den Hörgeräteträger mit entsprechendem Hördefizit.

Gemäß einer sinnvollen Weiterbildung des erfindungsgemäßen Verfahrens werden in einem weiteren Verfahrensschritt Testsignale erzeugt, die verschiedenen Vokalen mit hoch liegenden zweiten Formanten entsprechen, wobei der Hörgeräteträger die Unterscheidbarkeit der Testsignale bewertet. Zweiformantige Vokale mit hoch liegenden zweiten Formanten (z. B. /Y,i,e/) werden oft nicht gut genug in die Restgehördynamik transformiert. Es fehlt zum Teil eine verwertbare Überschwelligkeit der artikulativ schwachen zweiten Formanten in einem brauchbaren Entfernungsbereich. Eine ggf. individuell mangelhaft angepasste Dynamikkennlinie im kritischen Spektralbereich und

fehlende Abstimmung einer Begrenzungsfunktion können zu gravierender Fehlanpassung führen. Zu starke Resonanzen in der Otoplastik können darüber hinaus Kategoriengrenzen verschieben. Die insgesamt resultierende mangelhafte instabile Merkmalsübertragung führt zu einer schlechten Unterscheidbarkeit der Vokale untereinander und auch zu Verwechselungen mit /u/. Durch die Erfindung wird direkt überprüfbar gemacht, ob eine spektrale Anhebung der Energien der zweiten und dritten Formanten der kritischen Vokale deren Wahrnehmbarkeit verbessern würde. Dies kann unmittelbar in eine entsprechende Einstellung der Parameter des Hörgerätes umgesetzt werden. Nach Maßgabe der Bewertung durch den Hörgeräteträger können entsprechend die Dynamikkennlinien in den spektralen Bereichen der hoch liegenden Formanten eingestellt werden.

Die Vokalenergien sind die Träger von Sprachrhythmus oder segmentaler Betonung. So genanntes Recruitment, d. h. abnormer Lautheitszuwachs bei sensorischen Schäden, verändert die natürliche Wahrnehmung von Betonung und Rhythmus. Bei starker Variation der Schwelle und der Dynamik als Funktion des spektralen Ortes ist auch die Übertragung von Rhythmus stark verzerrt und bedarf einer Transformation auf ein konstantes Wahrnehmungsmaß für konstant artikulierte Rhythmusstärke. Dies gelingt durch eine spektral entsprechend unterschiedliche Kennliniensteigung der Kompression der relevanten Vokalmerkmalssignale im Bereich der dominanten Rhythmusübertragung (von ca. 250 bis ca. 1400 Hz). Darüber ist die Übertragung von Pegelunterschieden weniger kritisch. Um dies zu erreichen, können gemäß der Erfindung unbetont-betont Paare von vokalartigen Testsignalen verwendet werden. Die vom Hörgeräteträger empfundene Rhythmusstärke sollte für die verwendeten Testsignale bis zu Frequenzen von ca. 1400 Hz etwa gleich sein. Bei Abweichungen sollte die Kennliniensteilheit im betroffenen Spektralbereich nachgestellt werden.

Die Übertragung der wesentlichen Sprachatome im Resthörbereich muss auch dann sichergestellt sein, wenn Störungen durch Umgebungslärm vorhanden sind. Daher ist die Sicherstellung einer ausreichend Unterdrückung von Umgebungsstörlärm gleichermaßen unverzichtbar. Um dies zu kontrollieren, können bei dem erfindungsgemäßen Verfahren gleichzeitig mit den

Testsignalen Störlärmsignale erzeugt werden. Gegebenenfalls können die Störlärmsignale aus einem nicht frontalen Bereich des Hörgeräteträgers erzeugt werden. Auf diese Art und Weise kann die Wirksamkeit der Richtwirkung des Hörgerätes überprüft werden. Durch spektrales Herausfiltern von Störlärm werden nachteiligerweise auch Teile des Nutzsignals verändert, da eine echte spektrale Trennung von Sprachmerkmalssignalen und Lärmsignalen ggf. nicht möglich ist. Dies hängt von dem individuellen Umgebungslärm ab, dem der jeweilige Hörgeräteträger ausgesetzt ist. Durch die Bewertung der Überlagerung von Testsignal mit Störlärmsignal durch den Hörgeräteträger kann ein individueller Kompromiss gefunden werden, der durch entsprechende Einstellung der Parameter des Hörgeräts einerseits eine gewisse Lärmunterdrückung gewährleistet und andererseits noch keine störende Verstümmelung der Sprachbestandteile hervorruft. Weiterhin ermöglicht es das erfindungsgemäße Verfahren, gezielt eine Reduzierung des Abstandes zwischen dem Mikrofon des Hörgerätes und dem Sprechermund bei gleichzeitiger Reduzierung der Verstärkung herbeizuführen, so dass der wirksame Störpegel abgesenkt wird. Es wird dabei festgestellt, ob es einen brauchbaren Abstandsbereich gibt, in dem alle Sprachmerkmale ungestört übertragen werden. Der Hörgeräteträger kann daraus lernen, welchen Abstand er bei Vorliegen entsprechenden Umgebungsstörlärms zu seinem Gesprächspartner einhalten muss, um eine optimale Verständlichkeit zu gewährleisten.

Weiterhin betrifft die Erfindung eine Vorrichtung zum Verarbeiten von akustischen Sprachsignalen, mit einer elektronischen Verarbeitungseinrichtung, wobei die Verarbeitungseinrichtung zum lautklassenspezifischen Verarbeiten der Sprachsignale eingerichtet ist und Mittel aufweist, mit denen eine zeitliche Verlängerung von schwach artikulierten Lauten durchführbar ist.

Die erfindungsgemäße Vorrichtung ermöglicht die Realisierung des oben beschriebenen Verfahrens, nach dem eine individuelle Hervorhebung von schwach artikulierten Lauten erfolgen kann, wobei diese Hervorhebung nicht auf einer Verstärkung der Laute, sondern auf einer zeitlichen Verlängerung derselben beruht. Dadurch werden die im Zusammenhang mit dem Stand der Technik beschriebenen Nachteile vermieden.

Vorteilhafterweise weist die Vorrichtung eine Filtereinrichtung auf, mittels der die Sprachsignale in oberhalb einer oberen Grenzfrequenz liegende, hochfrequente Frequenzbänder und in unterhalb der oberen Grenzfrequenz liegende Frequenzbänder aufspaltbar sind. Die hochfrequenten Frequenzbänder können anschließend mittels Frequenzbandverschiebeeinheiten in den Nutzhörbereich unterhalb der oberen Grenzfrequenz individuell verschoben werden. Die unterhalb der oberen Grenzfrequenz liegenden Frequenzbänder können mittels Filtereinheiten der Filtereinrichtung individuell mit einer Pre-Emphasis moduliert werden.

Die Vorrichtung weist zweckmäßigerweise einen Lautklassenselektor auf, mit dem einem Sprachsignal eine bestimmte Lautklasse zugeordnet werden kann. Auf Basis der jeweils vorgenommenen Zuordnung einer Lautklasse zu einem Sprachsignal erfolgt vorzugsweise die zeitliche Verlängerung der schwach artikulierten Laute. Mittels einer Kompressoreinheit der Vorrichtung sind die Frequenzbänder individuell komprimierbar, wobei auch diese Komprimierung in Abhängigkeit der jeweils einem Sprachsignal zugeordneten Lautklasse gesteuert werden.

Des Weiteren wird eine Vorrichtung zum Verarbeiten von akustischen Signalen vorgeschlagen, mit einer elektronischen Verarbeitungseinrichtung, wobei die Verarbeitungseinrichtung zum Ersetzen von schwach artikulierten Lauten durch ihnen entsprechende, synthetische Laute eingerichtet ist.

Außerdem betrifft die Erfindung einen Sprachsynthesizer mit Mitteln zur Erzeugung von zwei oder mehr Formantwellenformen jeweils durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit einer Umhüllenden-Funktion, Mitteln zur Addition der zwei oder mehr Formantwellenformen und Mitteln zur Überlagerung und Verkettung der addierten Formantwellenformen gemäß einer Pitchintervalllänge zu einem Sprachsignal. Der Sprachsynthesizer ist zur Durchführung des oben beschriebenen Syntheseverfahrens eingerichet, bei dem

- zwei oder mehr Formantwellenformen jeweils durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit einer Umhüllenden-Funktion erzeugt werden,

- die zwei oder mehr Formantwellenformen addiert werden und
- die addierten Formantwellenformen gemäß einer Pitch-intervalllänge und gemäß suprasegmentaler Verkettungsregeln zu einem suprasegmentalen Sprachsignal verkettet werden. Dabei werden die Quellensignale bei der Erzeugung der Formantwellenformen frequenzmoduliert.

Das oben beschriebene Verfahren zur Kontrolle der Anpassung eines Hörgerätes kann vom Hörgeräteträger selbst sehr einfach angewandt werden. Fachpersonal ist nicht zwingend erforderlich. Hierzu bedarf es lediglich einer geeigneten Anordnung, die einen Personalcomputer, ein mit dem Personalcomputer verbundenes Audio-Interface und wenigstens ein mit dem Audio-Interface (z.B. über einen Verstärker) verbundenen Lautsprecher umfasst. Ein entsprechendes Computerprogramm für den Personalcomputer ermöglicht es, das oben beschriebene Verfahren durchzuführen. Wichtig für die Reproduzierbarkeit der Wahrnehmung der Testsignale gemäß der Erfindung ist ein angenähert linearer Ausgabefrequenzgang der Anordnung. Preisgünstige aktive PC-Lautsprecher weisen in der Regel nicht akzeptable Schwankungen des Frequenzgangs auf, so dass eine elektronische Kompensation erforderlich ist. Die erforderliche Frequenz kann mit geringem Aufwand mittels eines durch Software implementierten Linearisierungsfilters erreicht werden. Hierzu kann z.B. ein FIR-Filter mit konstanter Gruppenlaufzeit verwendet werden. Als Referenz zur Kalibrierung des Liniearisierungsfilters kann ein mit dem Audio-Interface verbundenes Mikrofon verwendet werden. Das Mikrofon sollte einen möglichst linearen Frequenzgang haben. Es eignen sich handelsübliche hochlineare Elektret-Mikrofone. Zur Eichung der Lautstärkepegel des Testsignals kann ebenfalls das Elektret-Mikrofon benutzt werden, z. B. in Verbindung mit einer geeichten Signalquelle. Mit diesem Zusatzinstrumentarium können einfache Personalcomputer (z. B. auch Laptops) mit zugehörigen kompatiblen Lautsprechern zur Durchführung des erfindungsgemäßen Verfahrens verwendet werden. Es sollte darauf geachtet werden, dass die Lautsprecher ausreichend hohe Leistungen haben, damit genügend klirrarme Reserven für höhere Pegel der Testsignale zur Verfügung stehen.

Die verschiedenen Aspekte der Erfindung werden im Folgenden anhand der in den Zeichnungen gezeigten Ausführungsbeispiele näher erläutert. Dabei zeigen:

Figur 1: eine schematische Darstellung eines Ausführungsbeispiels einer erfindungsgemäßen Vorrichtung;

Figur 2: eine schematische Darstellung eines weiteren Ausführungsbeispiels einer erfindungsgemäßen Vorrichtung;

Figur 3: ein Ausführungsbeispiel für die Synthese des Ersatz-„s“ und des Ersatz-„ch“; und

Figur 4: ein Ausführungsbeispiel für die Synthese des Ersatz-„t“.

Figur 5: Darstellung der erfindungsgemäßen Frequenzmodulation des Quellensignals bei der Erzeugung einer Formant-wellenform.

Figur 6: Spektrum und Zeitsignal des Testsignals /u/ gemäß der Erfindung;

Figur 7: Testsignal /o1/;

Figur 8: Testsignal /o2/;

Figur 9: Testsignal /a/;

Figur 10: Testsignal /ö/;

Figur 11: Testsignal /ü/;

Figur 12: Testsignal /i/;

Figur 13: Testsignal /ä/;

Figur 14: Testsignal /e/;

Figur 15: Testsignal Vokal /a/ mit kontrastierender Betonung (Zeitsignal);

Figur 16: Testsignal /k-h/, Zeitsignal und Spektrum;

Figur 17: Testsignal /t-e/ (Zeitsignal);

Figur 18: Testsignal /e-t-e/ (Zeitsignal);

Figur 19: Testsignal /i-e/ (Zeitsignal);

Figur 20: Testsignal /sch-f-ch-s/ (Zeitsignal und Spektrum von /s/);

Figur 21: Schematische Darstellung der erfindungsgemäßen Anordnung zur Kontrolle der Anpassung eines Hörgeräts;

Figur 22: Anordnung mit Mikrofon;

Figur 23: Anordnung zur Kontrolle der Richtwirkung eines Hörgerätes.

Die in Figur 1 gezeigte Ausführungsform der Vorrichtung 1 weist eine Filtereinrichtung 2 auf, mittels der die eingehenden akustischen Sprachsignale 3 in oberhalb der oberen Grenzfrequenz liegende, hochfrequente Frequenzbänder FB1, FB2 und FB3 und in unterhalb der oberen Grenzfrequenz liegende Frequenzbänder FB4 und FB5 aufgespaltet werden. Der dargestellte obere Bereich 4 der Filtereinrichtung 2 dient der Verarbeitung von nicht zu verschiebenden, unterhalb der oberen Grenzfrequenz liegenden Frequenzbändern FB1, FB2 und FB3 der Sprachsignale 3, wohingegen der dargestellte untere Bereich 5 der Filtereinrichtung 2 die hochfrequenten Frequenzbänder FB4 und FB5 aus den eingehenden Sprachsignalen 3 herausfiltert, welche oberhalb der oberen Grenzfrequenz liegen und in den Nutzhörbereich unterhalb der oberen Grenzfrequenz verschoben werden sollen. Die Vorrichtung 1 weist weiter einen Pitch-Synchronisator 6 auf, welcher zur Synchronisation der Fensterung der Frequenzbandverschiebeeinheiten 7 unter Berücksichtigung der Phase der Umhüllenden der Sprachsignale 3 über die Steuerleitung 8 dient. Des Weiteren weist die Vorrichtung 1 einen Lautklassenselektor 9 auf, welcher einem empfangenen Sprachsignal 3 eine

vorgegebene Lautklasse zuordnet. Das Ergebnis dieser Zuordnung wird zur Steuerung weiterer Komponenten der Vorrichtung 1 über Steuerleitungen 10, 11 und 12 verwendet, die im Folgenden beschrieben werden.

Eine dieser Komponenten der Vorrichtung 1 ist ein Frequenzschiebemodul 13, welches in dieser Ausführungsform zwei programmierbare Frequenzbandverschiebeeinheiten 7 aufweist. Mittels dieser Frequenzbandverschiebeeinheiten 7 kann eine individuelle Anpassung der Vorrichtung 1 an die jeweiligen Anforderungen erfolgen. Die Frequenzbandverschiebeeinheiten 7 arbeiten vorzugsweise mit Abtast-Modifikation. Das Spektrum, das jede Frequenzbandverschiebeeinheit 7 erzeugt, wird durch ein nachgeschaltetes Postfilter 14, 15 begrenzt. Diese sind als Bandpassfilter ausgebildet, welche das verschobene Signal im spektralen Bereich begrenzen und ein Überschreiten physiologischer Lautklassengrenzen verhindern. Das Ausgangssignal der Postfilter 14, 15 wird jeweils durch einen Kombinator 16, aufweisend eine angepassten Soft-Switch, durchgeschaltet oder abgeschaltet. Dieses Durch- oder Abschalten wird durch den Lautklassenselektor 9 in Abhängigkeit der jeweils einem Sprachsignal 3 zugeordneten Lautklasse über die Steuerleitung 10 gesteuert.

Die Vorrichtung 1 weist weiter ein Mittel 17 zum lautklassenspezifischen zeitlichen Verlängern von schwach artikulierten Lauten auf. Dieses tastet den Signalstrom, der aus dem Verknüpfer 18 zu ihm gelangt, mit überlappenden Fenstern ab, speichert den Fensterinhalt ab und gibt ihn nach vorgebbarer Zeit, etwa zwischen 2 und 10 ms, wieder aus und addiert ihn zum Eingangssignalstrom. Die Operationen „Verzögern" und „Addition" können parallel mit mehreren Verschiebe-Hall-Zeiten erfolgen. Im gezeigten Ausführungsbeispiel wird einmal um 4 ms verzögert und jeweils addiert. Für verschiedene Lautklassen können verschiedene Arbeitsweisen angewandt werden. Auch dieses wird durch den Lautklassenselektor 9 über die Steuerleitung 11 gesteuert.

Die Kompressoreinheit 19 ist ein 3-Band-Kompressor mit einer Kompressoreinheit K1, K2 beziehungsweise K3 und mit drei unterschiedlichen Zeitkonstanten je Band, wobei jeweils eine Zeitkonstante pro Bandlage an die Sprachmerkmale angepasst sind. Es gibt eine in allen Bändern identische

langsame Verstärkungseinstellung, eine mittel-schnelle Silbenkompression und eine schnelle Limitierung mit unterschiedlichen Geschwindigkeitscharakteristiken. Alle arbeiten mit „look-ahead" Technik und vermeiden so Einschwingspitzen. Eine geschwächte Rückwärtskopplung des zweiten auf das erste Band und des dritten auf das zweite Band wirkt dem physiologischen „upward spread of masking" entgegen. Im Frequenzbereich oberhalb von 1,5 kHz wird speziell durch die Art der Steuerung der schnellsten Kompressionskomponente eine Reduzierung des Spitzen-zu-Mittelwert Verhältnisses bei Formanten erreicht, so dass die Stärke der Formantinformation relativ zur Stärke der Stimmtonhöheninformation anwachsen kann und so für eine klarere Wahrnehmung des Vokalcharakters sorgen kann.

Über die Regeleinrichtung 20 erfolgt eine Vorverstärkung der zu komprimierenden Frequenzbänder vor ihrer Kompression, wozu die Kompressoreinheiten K1, K2 und K3 individuell von der Regeleinrichtung 20 über die Steuerleitungen 21 geregelt werden. Die Regeleinrichtung 20 selber wird über die Steuerleitung 12 in Abhängigkeit von der jeweils einem Sprachsignal 3 mittels des Lautklassenselektors 9 zugeordneten Lautklasse gesteuert. Die Ausgangssignale der einzelnen Kompressoreinheiten K1, K2 und K3 werden mittels des Verknüpfers 22 miteinander kombiniert und einem Equalizer 23 zugeführt. Dieser erzeugt das Ausgangssprachsignal 24 der Vorrichtung 1.

Figur 2 zeigt schematisch ein weiteres Ausführungsbeispiel für eine erfindungsgemäße Vorrichtung 201. Diese weist eine Verarbeitungseinheit 202 zum Verarbeiten der eingehenden Sprachsignale 203 im Resthörbereich auf. Diese Verarbeitungseinrichtung 2022 weist mehrere Kompressoreinheiten mit verschiedenen Kompressionscharakteristiken auf, um die eingehenden Sprachsignale 203 individuell an den jeweilig vorliegenden Hörschaden angepasst verarbeiten zu können. Aus den eingehenden Sprachsignalen 203 werden mittels eines Merkmalsextraktors 204 Sprachsignalmerkmale herausgefiltert. Die extrahierten Sprachsignalmerkmale werden dann an den Klassifikator 205 abgegeben, mit dem Lautprototypen, welche in einer Trainingsdatenbank 206 gespeichert sind, den Sprachsignalen 203 zugeordnet werden. Wird hierbei ein schwach artikulierter Laut erkannt, sind über einen Soft-Switch dem schwach artikulierten Laut entsprechende, synthetische Laute

den mittels der Verarbeitungseinrichtung 202 verarbeiteten Sprachsignalen über den Verknüpfer 208 zufügbar. Die synthetischen Laute werden in einer Synthese-Einrichtung 208 generiert und anschließend mittels einer Bearbeitungseinrichtung 210 komprimiert und moduliert. Die Kompression und Modulation erfolgt in Abhängigkeit von den erkannten Sprachsignalmerkmalen, indem der Merkmalsextraktor 204 die Bearbeitungseinrichtung 210 über die Steuerleitung 211 entsprechend steuert. Da die Erkennung der Sprachsignalsmerkmale, die Zuordnung von Lautprototypen zu dem jeweiligen Sprachsignal sowie die Generierung und Bearbeitung der synthetischen Laute Verarbeitungszeiten bedürfen, die nicht der für die Verarbeitung der eingehenden Sprachsignale 203 im Resthörbereich mittels der Verarbeitungseinrichtung 202 entsprechen, folgt eine Synchronisation der verarbeiteten Signale vor der Verknüpfung mittels des Verknüpfers 208 über eine Synchronisierungseinrichtung 212.

Figur 3 zeigt schematisch ein Ausführungsbeispiel für die Synthese des Ersatz-„s“ und des Ersatz-„ch“. Die eingehenden Sprachsignale 213 werden aufgespalten und im oberen dargestellten Zweig zur Erzeugung eines frequenzmodulierten Sinussignals und im unteren Zweig zur Erzeugung eines Rauschsignals verwendet. Im oberen Zweig durchläuft das Sprachsignal 213 zunächst ein Bandpassfilter 214, welches einen lautspezifischen Durchgangsbereich aufweist. Das bandpassgefilterte Sprachsignal wird anschließend einem Nulldurchgangszähler 215 zugeführt, um die momentane Frequenz aus der Zählung von Nulldurchgängen zu erhalten. Diese momentane Frequenz dient der Bestimmung des Frequenzschwerpunktes des Sprachsignals, welcher zur Modulation des Ersatzlautes bzw. zur idealen Anpassung desselben an den zu ersetzenden, schwach artikulierten Laut dient. Nach dem Nulldurchgangszähler 215 wird das Sprachsignal zu einem Sinusgenerator 216 geführt, mit welchem das gewünschte dem Rauschsignal zu überlagernde Sinussignal erzeugt wird. Dieses Sinussignal wird anschließend mit dem Rauschsignal des Rauschgenerators 217 über einen Verknüpfer 218 verknüpft. Der untere Zweig verwendet zunächst einen Bandpassfilter 219, mittels welchem ein lautspezifischer Frequenzbereich aus dem Sprachsignal 213 herausgefiltert wird. Dieses bandpassgefilterte Sprachsignal wird anschließend einer Einrichtung 220 zugeführt, die den Moving-Average bildet. Die resultierende

Originalenergie wird anschließend durch eine Kompressions-Kennlinie 221 bzw. 222 entsprechend der neuen spektralen Lage transformiert. Die transformierten Sprachsignale werden anschließend im unteren Zweig mit dem Rauschsignal des Rauschgenerators 223 über den Verknüpfer 224 verknüpft. Des Weiteren wird das transformierte Sprachsignal der Kompressions-Kennlinie 221 über den Verknüpfer 225 mit dem mittels des Verknüpfers 218 verknüpften Sprachsignals verknüpft. Mit dem weiteren Verknüpfer 226 wird das vom Verknüpfer 224 erzeugte Sprachsignal mit dem vom Verknüpfer 225 erzeugten Sprachsignal verknüpft und anschließend einem Soft-Switch 227 zugeführt, welcher dem Soft-Switch 207 der Figur 2 entspricht. Der Soft-Switch 227 wird Sprachsignalmerkmal-abhängig über den Sprachmerkmalsextraktor 204 der Figur 2 und die Steuerleitung 211 derart gesteuert, dass beim Auftreten eines schwach artikulierten Lautes ein Ersetzen desselben durch einen ihm entsprechenden, synthetischen Laut erfolgt.

Figur 4 zeigt ein schematisches Ausführungsbeispiel für die Synthese des Ersatz-„t". Hierzu werden eingehende Sprachsignale 228 mit einem Bandpass-filter 229, der einen Durchlassbereich zwischen 100 und 800 kHz hat, gefiltert. Das gefilterte Sprachsignal wird einer Einrichtung 230 zur Bildung des Moving-Average zugeführt. Das von dieser Einrichtung 230 ausgehende Sprachsignal wird aufgespalten und Kompressions-Kennlinien 231 und 232 zur Transformation der resultierenden Originalenergie entsprechend der neuen spektralen Lage zugeführt. Das von der Kompressions-Kennlinie 232 verarbeitete Sprachsignal wird über einen Verknüpfer 233 mit dem Rauschsignal des Rauschgenerators 234 verknüpft. Das Sprachsignal der oberen Kompressions-Kennlinie 231 wird ebenfalls einem Verknüpfer 235 zugeführt, welcher diese Sprachsignal mit in der Einrichtung 236 gespeicherten Plosionen verknüpft. Die in den Verknüpfern 233 und 235 verknüpften Sprachsignale werden mittels des Verknüpfers 238 miteinander verknüpft und einem Soft-Switch 239 entsprechend dem in Figur 2 dargestellten Soft-Switch 207 zugeführt, welcher ein Signal in Abhängigkeit von der Steuerung über die Steuerleitung 211 abgibt.

Figur 5, die das erfindungsgemäße Sprachsyntheseverfahren illustriert, zeigt im oberen Teil eine Umhüllende 301 einer Formantwellenform. Die

Formantwellenform wird durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit der Umhüllenden-Funktion 301 erzeugt. Rechts und links der Umhüllenden 301 sind gepunktet zeitlich vorangehende bzw. folgende Umhüllende 302, 303 weiterer Formantwellenformen des Sprachsignals dargestellt. Derartige verkettete und überlagerte Wellenformen bilden insgesamt das synthetisierte Sprachsignal. Die Formantwellenform besteht aus den zeitlich aufeinander folgenden Segmenten Einschwingsegment E, Haltesegment H und Ausschwingsegment A, welche gemäß dem oben beschriebenen Verfahren erzeugt werden, wobei das Ausschwingsegment A einer vorangehenden Formantwellenform das Einschwingsegment E der folgenden Formantwellenform, je nach Pitchintervalllänge, zeitlich überlappt.

Die beiden unteren Grafiken zeigen Ausführungsbeispiele für Funktionen, mit denen das Quellensignal bei der Erzeugung der Formantwellenform frequenzmoduliert, um das Auftreten von Tonalität zu verhindern. Es wird um die Mitte des Haltesegmentes H zentriert eine vollständige (obere Funktion) oder eine halbe (untere Funktion) Sinusschwingung verwendet, die im gezeigten Fall bis etwa y = 75 % in das Einschwingsegment E hineinragt. Der Modulationshub x beträgt in den Ausführungsbeispielen etwa 10 %.

Die Figuren 6 bis 20 zeigen beispielhaft Testsignale, die gemäß der Erfindung zur Kontrolle der Anpassung eines Hörgerätes verwendet werden können.

Wie oben ausgeführt, ist die individuelle Feinanpassung von Hörgeräten mit Sinussignalen, Schmalbandgeräuschen, Wortmaterial und Logatomen nicht geeignet, eine optimale Übertragung von Sprache zu prüfen oder einzustellen. Oft werden die artikulativ schwachen Sprachelemente nicht mit ausreichender Güte oder ausreichendem Pegel abgebildet.

Um diese Probleme zu beseitigen, wird gemäß der Erfindung ein Satz neuartiger Testsignale bereitgestellt. Diese Testsignale können aus Aufnahmen natürlicher Sprache gewonnen werden oder auch durch digitale Synthese.

Die erfindungsgemäßen Testsignale sind so ausgebildet, dass sie stets wie Elemente natürlicher Sprache wahrgenommen werden und entsprechend benennbar sind auch wenn sie nur aus spektralen Teilen derselben bestehen. Wesentliches Merkmal der Erfindung ist, dass die vokaläquivalenten Testsignale in solcher Weise ausgewählt oder spektral gefiltert sind, dass Adjustierungen der Filterbänke von Hörgeräten direkt vorgenommen werden können. Es kommt, anders ausgedrückt, darauf an, dass als Testsignale natürliche oder naturähnliche Signale mit spektral konzentrierten Merkmalsenergien verwendet werden.

Bei stark eingeschränktem Dynamikbereich des Gehörs (bei sensorischen Schwerhörigkeiten) besteht das Grundproblem, Verstärkungen und Kompressionskennlinien in den einzelnen spektralen Bändern des Hörgeräts so einzustellen, dass zunächst sämtliche Merkmale vokalischer Laute mit angenehmer Lautheit und bester Unterscheidbarkeiten neuronal abgebildet werden. Es soll ein bestwirksames vokalisches Exzitationsprofil erreicht werden. Bei Verwendung von natürlichem Sprachmaterial mit spektral ausgebreiteten multiplen Merkmalen ist es unmöglich, einzelne Spektralbereiche separiert einzustellen. Daher werden gemäß der Erfindung neuartige Testsignale zur Verfügung gestellt, die eine hinreichende spektrale Konzentriertheit lokal aufweisen und trotzdem Bausteine von Sprachelementen sind.

Es können Filterregeln aufgestellt werden, um erfindungsgemäß geeignete vokalähnliche Testsignale zu erzeugen. Hierzu wird z. B. der Spektralbereich in fünf Unterbereiche eingeteilt. 250 bis 400 Hz, 400 bis 600 Hz, 700 bis 1400 Hz, 1400 bis 2000 Hz und 2000 bis 3500 Hz. In diesen Bereichen sind die Formanten der Vokale in unterschiedlicher Weise auszufiltern, um übliche Anpassungsfehler direkt identifizieren oder vermeiden zu können.

Bei tiefliegenden Formanten < 400 Hz (/i, ü, u/), deren Pegel um mehr als 6 dB unter Formantpegeln des /o/s (400 bis 600 Hz) liegt, gilt, dass es einer relativen Verstärkungsanhebung bedarf, um auf gleiches Lautheitsniveau zu kommen. Andererseits darf die Exzitation in diesem Bereich nicht so hoch sein, dass Störgeräusche zu starke spektral aufwärtsgerichtete Maskierungen hervorrufen. Umgekehrt besitzt die Lautgruppe der /a/s relativ zu /o/ wesentliche höhere

Artikulationspegel (ca. 9 dB), so dass im Bereich 700 bis 1400 Hz eine Verstärkungsreduzierung sinnvoll ist. Keinesfalls darf eine Erhöhung der Exzitationspegel durch Verstärkungseinstellung oder akustische Schlauch-Hörer-Otoplastikresonanzen wirksam werden. Trotz heutiger verbesserter Resonanzdämpfung kann es durch mangelhafte Otoplastikdimensionierungen und unzureichende Dämpfungen solche Resonanzen geben. Den Bereich von 1400 bis 2000 Hz teilen sich die Merkmalsenergien des /ö/, des /ü/ und teilweise des /ä/. Der natürliche /ä/ Pegel liegt dabei um ca. 12 dB über dem Pegel des /ü/. Und der Pegel des /ü/ liegt um mehr als 15 dB unter dem des /o/. Um die Merkmalsenergie des /ü/ hörbar zu machen, bedarf es einer entsprechenden Überhöhung der Grundverstärkung in diesem Bereich. Dabei würde bei eingeengtem Dynamikbereich der zweite Formant des /ä/ bei linear steigender Kennlinie ohne Begrenzung in Bereiche jenseits der Unbehaglichkeitsgrenze verschoben. Um dem entgegenzuwirken, ist eine zweigeteilte Kennlinie mit geeignet ansteigendem eigentlichen Durchlassbereich und anschließender Begrenzung unbedingt erforderlich. Je nach Resthörbereichsweite (vom Pegel angenehmen Hörens bis zur Unbehaglichkeitsgrenze) ist es unvermeidlich, den Arbeitspunkt für den zweiten Formanten des /ä/ mehr oder weniger in den Begrenzungsbereich zu schieben. Eine akzeptable /ä/ Lautheit muss sehr präzise eingestellt werden. Keinerlei Resonanzüberhöhungen im Frequenzgang des Gerätes sind bei geringen Dynamikumfängen erlaubt, da diese unter Umständen zu effektiven Verschiebungen des Wirkungsschwerpunktes des Komplexes aus zweiten und dritten Formanten (z. B. bei /ü/ und /i/) führen könnten und damit zu Lautverwechselungen. Den Bereich zwischen 2000 und 3500 Hz teilen sich /i/ und /e/ mit der oberen Spektralhälfte von /ä/, hierbei weist das /i/ ähnlich niedrige Spektralpegel auf wie das /ü/ im vorhergehenden Bereich. Entsprechend hoch muss die Grundverstärkung für den zweiten und dritten Formanten des /i/ sein, um für die auditive Sprachaufnahme verwertbar zu sein. Entsprechend der zur Verfügung stehenden Dynamik muss die Begrenzung des /ä/ eingestellt werden.

Es ist in der Regel bei sensorischen Hörschäden unbekannt, in welcher Weise spektrale Energien zu Summenlautheiten integriert werden. Also muss die Sicherstellung von Mindestlautheiten bei schwachen Merkmalen und die Sicherstellung der Begrenzung von Maximallautheiten (um weitere

Schädigungen des Gehörs zu vermeiden) bei starken Merkmalen durch das gewählte Einstellverfahren und durch die geeignete Wahl der Testsignale gewährleistet sein. Entsprechend sind die vokaläquivalenten Testsignale der Figuren 6 bis 14 jeweils aus dem Gesamtspektrum der natürlichen Laute herausgefiltert, und zwar unter Anwendung von phasenlinearen FIR Filtern.

Die Figur 6a zeigt das Zeitsignal und die Figur 6b das Spektrum des Testsignal /u/. Im niederfrequenten Bereich bis zu 1000 Hz muss darauf beachtet werden, dass oft starke Störgeräusche vorhanden sind, mit Anregungsfrequenzen unterhalb von 200 Hz und starkem Oberwellenspektrum. Daher ist grundsätzlich eine niedrigst mögliche Verstärkung zu wählen, so dass die Störgeräusche soweit wie möglich keinen Einfluss auf die Sprachwahrnehmung im Bereich > 1000 Hz haben. Daraus folgt, dass die Übertragung des /u/ so eingestellt werden muss, dass bei geringst möglicher Grundverstärkung ein leises /u/ noch gerade gut wahrgenommen wird. Für die erfindungsgemäße Kontrolle der Anpassung des Hörgerätes ist die Wirkung des /u/ sehr wichtig. Sämtliche Komponenten des /u/, die wahrnehmungsrelevant sein können, müssen zur Verfügung stehen. Bei dem in der Figur 1 dargestellten Testsignal ist das natürliche Spektrum des /u/ zwischen 250 und 500 Hz bandpassgefiltert.

Die Figur 7a zeigt das Spektrum und die Figur 7b das Zeitsignal des Testsignals /o1/. Es handelt sich um das offene /o/, das, wie das /u/, noch im Bereich von niederfrequenten Lärmkomponenten liegt und in einem Bereich, in dem hohe Pegel stark aufwärtsmaskierend wirken können. Trotz möglicherweise großem Dynamikbereich muss dafür gesorgt werden, dass keine zu hohen Lautheiten auftreten. Die Kennliniensteilheit muss rhytmusoptimal abbilden und die Horizontalbegrenzung muss eher konservativ bei der Bewertung des Testsignals durch den Hörgeräteträger als laut erfolgen. Der erste Formant des offenen /o/ wird zur Erzeugung des in der Figur 7 gezeigten Testsignals breitbandig zwischen 250 und 700 Hz herausgefiltert.

Die Figur 8 zeigt in entsprechender Weise das Testsignal /o2/. Es handelt sich um das geschlossene /o/. Es gilt dasselbe für das offene /o/. Entsprechend der natürlichen Bandbreite wird der erste Formant breitbandig zwischen 300 und 900 Hz herausgefiltert.

Die Figur 9 zeigt das Testsignal /a/. Alle Varianten des /a/ weisen relativ zu den benachbarten Vokalenergien sehr viel höhere Pegel auf. Daher ist die Gefahr gegeben, dass die /a/-Exzitation zu zu hoher Lautheit führt. Damit eine solche Überexzitation, die als solche keinen Merkmalscharakter besitzt und die schwache zweite Formanten in der spektralen Umgebung maskieren kann, vermieden wird, müssen Dynamikkennlinie und Grundverstärkung unter Benutzung des vollständigen Haupt-/a/-Merkmalskomplexes eingestellt werden, damit die volle Lautheit-Summation bei der Kontrolle wirksam werden kann. Daher wird das Testsignal /a/ gemäß Figur 9 im Bereich zwischen 600 Hz und 1600 Hz mit beiden Hauptformanten extrem breitbandig ausgefiltert. Wenn die Lautheit dieses Komplexes dynamisch isophon relativ zu den Merkmalen der anderen Vokale richtig einjustiert wird, kann man davon ausgehen, dass eine exzessive Verdeckung verhindert wird.

Die Figur 10 zeigt das Testsignal /ö/. Die relativ schwache Merkmalsenergie des /ö/ liegt am Ende des /a/-Spektralbereiches und kann daher durch entsprechende Verstärkung angehoben werden. Entsprechend der natürlichen Breite des zweiten Formanten wird die Energie zwischen 1100 bis 1800 Hz herausgefiltert, wie in Figur 10 gezeigt.

In Figur 11 ist das Testsignal /ü/ abgebildet. Die Merkmalsenergie des /ü/ ist extrem schwach und schmalbandig. Auf gute Überschwelligkeit muss durch Einstellung einer geeigneten Verstärkung in diesem Spektralbereich besonders geachtet werden. Bei Unterschwelligkeit wird das /ü/ zu /u/. Entsprechend ist zur Erzeugung des Testsignals gemäß Figur 11 die Filterbreite gewählt zu 1750 bis 2100 Hz.

Die Figur 12 zeigt das Testsignal /i/. Das /i/ kann im Pegel noch niedriger liegen als das /ü/ und benötigt daher noch mehr Basisverstärkung. Im Unterschied zum /ü/ besitzt das /i/ allerdings nicht nur einen einzigen sondern zwei höhere merkmalstragende Formanten, die bei Lautheitssummation durch verbreiterte kritische Bänder sich durchaus verstärken können. Beide müssen also bei der Kontrolle der Anpassung eines Hörgerätes gemäß der Erfindung berücksichtigt werden. Die Bandfilterung des Testsignals /i/ gemäß Figur 12 erfolgt daher im Bereich von 2050 bis 3300 Hz.

Probleme bereitet das /ä/. Das entsprechende Testsignal ist in Figur 13 gezeigt. Die Merkmalsenergie des /ä/ wird im Bereich zwischen 1000 Hz und 2600 Hz für das Testsignal herausgefiltert, um sämtliche spektralen Anteile, die um die Lage des Maximums herum verteilt sind und die lautheitserzeugend sind, bei der Kontrolle berücksichtigen zu können. Nur so kann die Limitierung der starken /ä/-Energien durch geeignete Wahl der horizontalen Äste der entsprechenden Dynamikkennlinien in diesem Bereich adäquat einjustiert werden.

Die Figur 14 schließlich zeigt das Spektrum und das Zeitsignal des Testsignals /e/. Das /e/ besitzt Merkmalsenergie im Bereich von 1900 bis 2600 Hz und wird entsprechend ausgeschnitten. Nach vorheriger Justierung des /i/ und /ä/ fällt das /e/ automatisch in einen adäquaten Zwischenbereich der Dynamikkennlinie. Feinheiten der Steigung im Hauptdurchlassbereich können dennoch angepasst werden.

Mit den in den Figuren 6 bis 14 gezeigten bandpassartig ausgeschnittenen Vokalen als Testsignale lässt sich eine valide Grundeinstellung eines Hörgerätes mit ausreichender spektraler Verstärkung, Kompressionskennliniensteigungen und Begrenzungen erzielen.

Um eine hörrichtige Abbildung von „leise-laut"-Intonation zu gewährleisten, werden gemäß der Erfindung sinnvollerweise gleiche Testsignale wiederholt mit unterschiedlichen Lautstärken erzeugt, wobei nach Maßgabe der Bewertung durch den Hörgeräteträger die Kennlinien der spektral selektiven Dynamikkompression eingestellt werden. Vokalformanten sind, wie oben ausgeführt, Träger suprasegmentaler Informationen. Diese ist unverzichtbar zur neuronalen Segmentation von Sprache und muss daher sorgfältig erhalten bleiben. Die Wahrnehmbarkeit von Rhythmusunterschieden ist umso wichtiger, je stärker die Frequenzauflösung bei sensorischen Hördefiziten zerstört ist. Das bedeutet, dass eine den Rhythmus, bzw. die Pegelunterschiede aufgrund von Betonung erhaltende Dynamiktransformation via Kennliniensteilheit einstellbar sein muss. Das gilt hauptsächlich für Komponenten im Bereich zwischen 250 und 1500 Hz. Höher liegende Formanten sollten komoduliert mit den unteren Formanten sein zwecks besserer perzeptiver Fusion. Arbeitspunkte in begrenzenden Kennlinienästen sollten nur im Ausnahmefall (z. B. bei /ä/) zugelassen werden. Um eine

präzise Einstellung einer optimalen Sprachrhythmusübertragung zu gewährleisten, wird die zuvor beschriebene spektrale Filterung von natürlichen Sprachelementen zur Erzeugung der Testsignale auf natürliche (oder synthetische) Sprachrhythmus-Paare angewendet. Die Paare enthalten einen Pegelsprung von ca. 6 dB, dessen Wahrnehmung durch geeignete Wahl der Kompressionssteilheit im jeweils beaufschlagten Frequenzband derartig einjustiert werden muss, dass alle vokalartigen Testsignale in ähnlicher Weise „unbetont-betont" bzw. „leise-laut" vom Hörgeräteträger wahrzunehmen sind. Die bei natürlicher Intonation kovariierende Erhöhung der Stimmtongrundfrequenz wird unterdrückt, damit nicht irrtümlich eine Stimmtonänderung für eine Pegelerhöhung gehalten werden kann. Die Figur 15 zeigt beispielhaft das entsprechende Rhythmuspaar des Testsignals /a/. Die ersten vier Testsignale (Figuren 6 bis 9) sollten eine sehr ähnliche und deutliche Rhythmusstärke produzieren. Die folgenden fünf Signale (Figuren 10 bis 14) sollten wenigstens einen noch gerade wahrnehmbaren Rhythmus produzieren.

Die in den weiteren Figuren 16 bis 20 gezeigten Testsignale können bei dem erfindungsgemäßen Verfahren zusätzlich verwendet werden, zum Zwecke einer weitergehenden Verfeinerung der Anpassung. Es geht dabei um die Evaluierung der Wahrnehmbarkeit von Plosivmerkmalen und Frikativen. Es können einfache Aufnahmen von natürlicher Sprache hierzu verwendet werden.

Die Figur 16a zeigt das Zeitsignal und die Figur 16b das zugehörige Spektrum des Testsignals /k-h/. Es handelt sich um einen Plosivburst mit Aspiration. Dies sollte vom Hörgeräteträger gut wahrnehmbar sein. Die genauen Mechanismen für die spektrale und temporale Energiesummation der spektral breit ausgedehnten und temporal eng begrenzten Burstenergie sind weitgehend unbekannt. Die Empfindlichkeiten des geschädigten Gehörs können nicht aus Schwellenmessungen gefolgert werden. Daher ist es geboten, mit repräsentativen Prototypen direkt die Wahrnehmung festzustellen. Hierzu können Testsignale aus überdeutlich artikulierten Sprechproben der Plosiv-Bursts /p/, /t/ und /k/ mit Aspiration im stimmlosen /h/ Kontext präpariert werden. Alternativ können synthetisch erzeugte Prototypen verwendet werden. Zur Kontrolle der Anpassung eines Hörgeräts sollte untersucht werden, ob die Wahrnehmungs-

schwellen bei den einzelnen Testsignalen gut überschritten werden. Weiterhin sollte die Unterscheidbarkeit der Prototypen geprüft werden.

Um die noch zusätzlich vorhandene Wirkung von Formantübergängen, die im Kontext mit stimmhaften Vokalen merkmalsgebend ist, prüfen zu können, sind Plosiv-Vokal-Logatome als Testsignale erforderlich. Es sollten Kombinationen mit Übergängen in allen spektralen Bereichen geprüft werden. Besonders kritisch sind jedoch Kombinationen mit hoch liegenden zweiten Formanten, also mit /ü/, /e/, /i/, /ä/. Stellvertretend zeigt die Figur 17 das Zeitsignal des Testsignals /t-e/. Anhand dieses Testsignals kann die Übertragung speziell im kritischen Bereich > 2000 Hz fein justiert werden. Ergänzend können zur Prüfung der Verdeckung von Folgeplosiven durch vorangehende Vokale auch die Logatome mit der invertierten Reihenfolge als Testsignale zur Verfügung gestellt werden, z. B. /e-p/, /e-k/, /e-t/.

Bei überdeutlich artikulierter Sprachproduktion können Plosive mit überhöhtem Artikulationsdruck produziert werden. Dadurch kann die Wahrnehmbarkeit der Plosive für Personen mit Hördefiziten wesentlich gesteigert werden. Um zu überprüfen, ob dem Hörgeräteträger mit überdeutlich artikulierter Sprache geholfen werden kann, können als Testsignale bei dem erfindungsgemäßen Verfahren Signale verwendet werden, bei denen der Plosiv zwischen zwei Vokalen eingebettet ist, wobei der erste Vokal unbetont und der zweite betont ist. Die Figur 18 zeigt das Zeitsignal des entsprechenden Testsignals /e-t-e/ (mit Betonung auf dem zweiten /e/). Alternative Testsignale sind z.B. /e-p-e/ und /e-k-e/.

Bei Hochtondefiziten wird oft der Frequenzbereich > 2000 Hz nicht in ausreichender Qualität übertragen. Die Justierung der Exzitationsenergien von schwachen hoch liegenden zweiten Formanten relativ zu den Exzitationsenergien der ersten Formanten und den Stimmtonhöhen-Oberwellen kann mangelhaft sein. Als Folge ist das Hörbild, das erst aus der Kombination der Wirkung von erstem und zweitem Formanten entsteht, unscharf, oder die Wirkung des zweiten Formanten ist gar nicht vorhanden, so dass die Wahrnehmungsbilder in das Bild des /u/ kollabieren. Im ersten Schritt wurde, wie oben erläutert, die gute Wahrnehmung der Einzelformanten im Spektrum

sichergestellt. Eine Feinjustierung des Verhältnisses der Energien beider Formanten, welche die Simultanbeeinflussung der beiden Energien berücksichtigt, ist zur Einstellung oder Prüfung der am besten kognitiv klassifizierbaren Hörbilder unabdingbar. Dazu wird ergänzend ein /u-i-y-e/ Unterscheidbarkeitstest zur Verfügung gestellt. Die Figur 19 zeigt beispielhaft das Zeitsignal des Testsignals /i-e/. Weitere mögliche Testsignale sind /i-ü/, /u-ü/, /i-u/ sowie auch /u-i-ü-e/.

Ergänzend können bei dem erfindungsgemäßen Verfahren Testsignale für die Überprüfung der Wahrnehmung von Frikativen verwendet werden. Die Wahrnehmung von Frikativenergien, die naturgemäß in höheren spektralen Bereichen merkmalstragend sind, ist bei sensorischen Hochtondefiziten schwer bis sehr schwer gestört. Das kann dazu führen, dass mit normalen, nur verstärkenden Hörgeräten das /s/ und /h/ nur so schwach wahrgenommen werden, dass diese Laute praktisch für die Wahrnehmung laufender Sprache nicht nutzbar sind. Es müssen dann transformierte bzw. spektral transponierte Merkmalsenergien, die natürliche Merkmalsenergie ersetzen (sogenannte Transposer-Hörgeräte), zur Verfügung gestellt werden. Das /sch/ liegt in einem Bereich niedrigerer Frequenzen und wird im Vergleich zu den anderen Frikativen mit höchstem Artikulationspegel produziert. Hier kann oft durch ausreichende Verstärkung eine gute Auswertbarkeit erreicht werden. Das /f/ hat auch im Bereich niedriger Frequenzen merkmalstragende Energien. Diese haben allerdings sehr niedrige Pegel, so dass exzessive Verstärkung benötigt würde.

Zur Überprüfung der Wahrnehmbarkeit der stimmlosen Frikative kann das in der Figur 20 gezeigte Testsignal /sch-f-ch-s/ verwendet werden. Die Figur 20a zeigt das Zeitsignal, die Figur 20b das Spektrum des Bestandteils /s/.

Zur Ergänzung kann auch die Überprüfung der Wahrnehmbarkeit der stimmhaften Frikative erfolgen. Hier kann im Einzelfall die Ballance zwischen dem niederfrequenten stimmhaften Anteil und dem hochfrequenten stimmlosen Anteil der stimmhaften Frikative für eine gute Unterscheidbarkeit wichtig sein. Der stimmhafte Anteil darf den stimmlosen Anteil nicht verdecken. Um dies prüfen zu können, wird zweckmäßigerweise zusätzlich die Reihe der stimm-

haften Frikative /dj-w-j-z/ aufgenommen und ohne Filterung als Testsignal verwendet.

Die Figuren 21 bis 23 zeigen eine Anordnung zur Kontrolle der Anpassung eines Hörgerätes gemäß der Erfindung. Die Anordnung umfasst einen Personalcomputer 401 (ein Laptop), das, zum Beispiel über eine USB-Schnittstelle, mit einem Audio-Interface 402 an sich üblicher Art verbunden ist. An den Ausgang des Audio-Interfaces 402 ist ein Verstärker 403 mit Bedienelement 404 zur Verstärkungseinstellung angeschlossen. Mit dem Verstärker 403 ist ein Lautsprecher 405 verbunden. Der Lautsprecher 405 befinden sich frontal vor einem Hörtgeräteträger 406. Dieser trägt ein Hörgerät 407 hinter dem Ohr. Optional kann ein Absorptionstrichter 408 verwendet werden, der aus handelsüblichen Akustik-Dämmmatten besteht. Der Hörgeräteträger 406 befindet sich in einem Abstand von vorzugsweise etwa 1 Meter vor dem Lautsprecher 405. Auf dem Laptop 401 ist das oben beschriebene erfindungsgemäße Verfahren mittels entsprechender Software implementiert. Die Software kann vom Hörgeräteträger 406 selbst bedient werden, so dass die Durchführung des erfindungsgemäßen Verfahrens keines weiteren Fachpersonals bedarf.

Die Figur 22 zeigt die Anordnung mit zusätzlichem Mikrofon 409. Es handelt sich um ein hochlineares Elektret-Mikrophon, das zur Kalibrierung eines durch Software in dem Personalcomputer 401 realisierten Linearisierungsfilters dient. Das Linearisierungsfilter ist erforderlich, um den Frequenzgang des Lautsprechers bei der Ausgabe der Testsignale zu linearisieren.

Bei dem in Figur 23 in Draufsicht dargestellten Ausführungsbeispiel befindet sich ein weiterer Lautsprecher 405' schräg hinter dem Hörgeräteträger 406. Der Lautsprecher 405' dient zur Erzeugung eines Störsignals, um die Richtwirkung des zu kontrollierenden Hörgerätes 407 zu überprüfen. Die Standardmessposition des Lautsprechers 405' bezüglich des Lautsprechers 405 beträgt 115°.

Anmerkung:

Um die Verständlichkeit dieses Textes zu gewährleisten, wird die internationale phonetische Schreibweise nicht verwendet, sondern es wird eine graphemische Umschreibung verwendet. Der entsprechende Buchstabe wird zwischen Schrägstrichen wiedergegeben (z.B. /o/ für o wie „Ochse“, /ch/ wie „China“, /w/ wie „Ware“ oder /z/ für das stimmhafte „s“ wie in „See“).

- Patentansprüche -

Patentansprüche

1. Verfahren zum Verarbeiten von akustischen Sprachsignalen mittels einer elektronischen Verarbeitungseinrichtung,
d a d u r c h g e k e n n z e i c h n e t, dass

- eine lautklassenspezifische Verarbeitung der Sprachsignale erfolgt, wobei schwach artikulierte Laute zeitlich verlängert werden, und/oder
- schwach artikulierte Laute erkannt und durch ihnen entsprechende, synthetische Laute ersetzt werden.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die zu verarbeitenden Sprachsignale in mehrere Frequenzbänder aufgeteilt werden.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass hochfrequente Frequenzbänder, die oberhalb einer oberen Grenzfrequenz liegen, zu niedrigeren Frequenzen unterhalb der oberen Grenzfrequenz und oberhalb einer unteren Grenzfrequenz verschoben werden.

4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, dass die Verschiebung der hochfrequenten Frequenzbänder individuell in Abhängigkeit von der jeweiligen Frequenzlage eines hochfrequenten Frequenzbandes erfolgt.

5. Verfahren nach Anspruch 3 oder 4, dadurch gekennzeichnet, dass die unterhalb der oberen Grenzfrequenz liegenden Frequenzbänder mit unterschiedlicher Pre-Emphasis moduliert werden.

6. Verfahren nach einem der Ansprüche 3 bis 5, dadurch gekennzeichnet, dass die unterhalb der oberen Grenzfrequenz liegenden Frequenzbänder individuell komprimiert werden.

7. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass den Sprachsignalen vorgegebene Lautklassen zugeordnet werden.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, dass die zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbänder in Abhängigkeit der ihnen jeweilig zugeordneten Lautklasse zu einem Zwischensprachsignal miteinander kombiniert werden.

9. Verfahren nach einem der Ansprüche 3 bis 8, dadurch gekennzeichnet, dass die zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbänder mit einem oberen Frequenzband, welches unterhalb der oberen Grenzfrequenz am nächsten zu dieser angeordnet ist, zu einem Hochtonzwischensprachsignal kombiniert werden.

10. Verfahren nach Anspruch 8 oder 9, dadurch gekennzeichnet, dass das Zwischensprachsignal beziehungsweise das Hochtonzwischensprachsignal in Abhängigkeit von der den Sprachsignalen jeweilig zugeordneten Lautklasse gespeichert, in vorgebbaren Zeitabständen abgerufen, individuell komprimiert und zur Erzeugung eines Ausgangssprachsignals mit den übrigen individuell komprimierten Frequenzbändern, die unterhalb der oberen Grenzfrequenz liegen, kombiniert wird.

11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, dass Ausgangssprachsignale zur Kompensation von spektralen Eigenschaften von an die Verarbeitungseinrichtung anschließbaren Sprachsignalausgabeeinheiten mittels eines Equalizers moduliert werden.

12. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass eine Vorverstärkung der Frequenzbänder vor ihrer

Kompression in Abhängigkeit von der den Sprachsignalen jeweilig zugeordneten Lautklasse und/oder der Lautstärke von Umgebungsgeräuschen erfolgt.

13. Verfahren nach einem der Ansprüche 3 bis 12, dadurch gekennzeichnet, dass die Frequenzverschiebungen ausschließlich über Intra-Laut-Transformationen erfolgen.

14. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die lautspezifische Verarbeitung der Sprachsignale für jede Lautklasse eigene oder wenigstens zwei Lautklassen übergreifende Verarbeitungsmaßnahmen umfasst.

15. Verfahren nach einem der Ansprüche 3 bis 14, dadurch gekennzeichnet, dass die Frequenzverschiebungen mittels individuell programmierbarer Frequenzverschieber erfolgt.

16. Verfahren nach einem der Ansprüche 3 bis 15, dadurch gekennzeichnet, dass das niedrigste zu niedrigen Frequenzen verschobene, hochfrequente Frequenzband immer Signale liefert, wohingegen die frequenzmäßig höher liegenden, zu niedrigen Frequenzen verschobenen, hochfrequenten Frequenzbänder lautklassenspezifisch zugeschaltet werden.

17. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die synthetischen Laute vor ihrem Einfügen in das Sprachsignal hinsichtlich der Energie und/oder des Frequenzschwerpunktes der zu ersetzenden, schwach artikulierten Laute modifiziert werden.

18. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Sprachsignal vor dem Einfügen der synthetischen Laute zeitlich verzögert wird.

19. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die synthetischen Laute dynamisch in das empfangene Sprachsignal ein- und ausgeblendet werden.

20. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die schwach artikulierten Laute einem vorgegebenen Lautprototyp zugeordnet werden.

21. Verfahren nach Anspruch 20, dadurch gekennzeichnet, dass die Zuordnung der schwach artikulierten Laute zu den Lautprototypen unter Berücksichtigung wenigstens eines Sprachsignalmerkmals erfolgt.

22. Verfahren nach Anspruch 21, dadurch gekennzeichnet, dass als Sprachsignalsmerkmal eine Pause im Sprachsignal, die Lautstärkeänderungen des Sprachsignals in der Zeit, die Stimmhaftigkeit des Sprachsignals und/oder die spektralen Energieverhältnisse des Sprachsignals erfasst werden.

23. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die synthetischen Laute erzeugt werden, indem eine Rauschsignalkomponente und eine Sinussignalkomponente für einen synthetischen Laut generiert und miteinander kombiniert werden.

24. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die Amplituden der synthetischen Laute vor dem Einfügen in das Sprachsignal individuell komprimiert werden.

25. Sprachsyntheseverfahren, insbesondere zur Erzeugung von synthetischen Lauten in einem Verfahren nach einem der vorhergehenden Ansprüche , wobei

- zwei oder mehr Formantwellenformen jeweils durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit einer Umhüllenden-Funktion erzeugt werden,
- die zwei oder mehr Formantwellenformen addiert werden und
- die addierten Formantwellenformen gemäß einer Pitch-

intervalllänge und gemäß suprasegmentaler Verkettungsregeln zu einem suprasegmentalen Sprachsignal verkettet werden,
d a d u r c h g e k e n n z e i c h n e t ,
dass die Quellensignale bei der Erzeugung der Formantwellenformen frequenzmoduliert werden.

26. Verfahren nach Anspruch 25, dadurch gekennzeichnet, dass die Frequenzmodulation der Quellensignale zyklostationär ist.

27. Verfahren nach Anspruch 25 oder 26, dadurch gekennzeichnet, dass die Verkettung der frequenzmodulierten Wellenpakete derart durch pitchadaptive Umhüllendenformung erfolgt, dass keine wahrnehmbaren Störungen durch Modulation im Überlagerungsbereich der Wellenzüge auftreten.

28. Verfahren nach einem der Ansprüche 25 bis 27, dadurch gekennzeichnet, dass der Modulationshub der Frequenzmodulation bei der Erzeugung der Formantwellenformen von der jeweiligen mittleren Formantfrequenz abhängt.

29. Verfahren nach Anspruch 28, dadurch gekennzeichnet, dass der Modulationshub der Frequenzmodulation bis zu 20 %, vorzugsweise bis zu 10 % der jeweiligen mittleren Formantfrequenz beträgt.

30. Verfahren nach einem der Ansprüche 25 bis 29, dadurch gekennzeichnet, dass der Modulationshub der Frequenzmodulation bei der Synthese weiblicher Sprache kleiner ist als bei der Synthese männlicher Sprache.

31. Verfahren nach einem der Ansprüche 25 bis 30, dadurch gekennzeichnet, das bei der Überlagerung und Verkettung der addierten Formantwellenformen die Pitchintervalllänge variiert wird.

32. Verfahren nach Anspruch 31, dadurch gekennzeichnet, dass die Pitchintervalllänge zufällig variiert wird.

33. Verfahren nach Anspruch 31 oder 32, dadurch gekennzeichnet, dass die Pitchintervalllänge so variiert wird, dass der stilisierte Momentanwert um vorzugsweise 1% bis 2% bei der Synthese männlicher Stimmen und bis zu 0.5% bei der Synthese weiblicher Stimmen vom lokalen Mittelwert abweicht.

34. Verfahren nach Anspruch 32 oder 33, dadurch gekennzeichnet, dass horizontale, d. h. konstante Pitch-Stilisierungssegmente vermieden werden und dass vorzugsweise bei der Synthese von fast konstantem Pitchverlauf über typische Silbensegmentlängen von etwa 200 ms eine Pitchsteigung von mindestens 3% gewählt wird.

35. Verfahren nach einem der Ansprüche 25 bis 34, dadurch gekennzeichnet, dass die Umhüllenden-Funktion aus drei zeitlich aufeinander folgenden Segmenten, nämlich einem Einschwingsegment, in dem die Amplitude des Quellensignals von Null ausgehend ansteigt, einem Haltesegment, in dem die Amplitude des Quellensignals konstant ist, und einem Ausschwingsegment, in dem die Amplitude des Quellensignals wieder auf Null abfällt, besteht.

36. Verfahren nach Anspruch 35, dadurch gekennzeichnet, dass die Dauer des Einschwingsegments, des Haltesegments und/oder des Ausschwingsegments von der Pitchintervalllänge abhängen.

37. Verfahren nach Anspruch 36, dadurch gekennzeichnet, dass bei sinkender Pitchintervalllänge zunächst die Dauer des Ausschwingsegments bis auf einen Minimalwert verkürzt wird und dann die Dauer des Haltesegments verkürzt wird, so dass Interferenzen von Formantwellenformen aufeinander folgender Pitchintervalle vermieden werden.

38. Verfahren nach einem der Ansprüche 35 bis 37, dadurch gekennzeichnet, dass die Dauer des Einschwingsegments einer ganzzahligen Anzahl von Nulldurchgängen der Oszillation des Quellensignals entspricht.

39. Verfahren nach einem der Ansprüche 35 bis 38, dadurch gekennzeichnet, dass das Ausschwingsegment der Umhüllenden-Funktion so ausgebildet ist, dass die Amplitude zum Ende des Pitchintervalls auf höchstens 35%, vorzugsweise auf höchstens 25 % der während des Haltesegments konstanten Amplitude abgefallen ist.

40. Verfahren nach einem der Ansprüche 25 bis 40, gekennzeichnet durch eine Hochpassfilterung des verketteten Sprachsignals.

41. Verfahren nach Anspruch 40, dadurch gekennzeichnet, dass die Hochpassfilterung mittels eines IIR-Filters mit einer Cutoff-Frequenz von 100 Hz erfolgt.

42. Verfahren nach einem der Ansprüche 25 bis 41, dadurch gekennzeichnet, dass bei Änderung der Pitchintervalllänge die Formantfrequenzen pitch-abhängig variiert werden, und zwar in der Weise, dass die Formantfrequenzen bei Verkürzung der Pitchintervalllänge erhöht werden.

43. Verfahren zur Kontrolle der Anpassung eines Hörgerätes, das eine Filterbank zur spektral selektiven Verstärkung und Dynamikkompression von Audiosignalen aufweist, an ein Hördefizit eines Hörgeräteträgers, wobei mittels einer Signalquelle ein Testsignal erzeugt wird und die Wahrnehmung des Testsignals durch den Hörgeräteträger bewertet wird,
d a d u r c h g e k e n n z e i c h n e t ,
dass das Testsignal wenigstens ein natürliches oder naturähnliches Sprachelement umfasst, das in der Weise spektral gefiltert oder ausgewählt ist, dass das Spektrum des Testsignals zu dem Spektralbereich wenigstens eines Filters der Filterbank des Hörgerätes korrespondiert.

44. Verfahren nach Anspruch 43, dadurch gekennzeichnet, dass eine Mehrzahl von Testsignalen erzeugt wird, die unterschiedliche Spektralbereiche der Filterbank abdecken, wobei die spektral selektive Verstärkung der Filterbank nach Maßgabe der Bewertung der Testsignale durch den Hörgeräteträger eingestellt wird.

45. Verfahren nach Anspruch 43 oder 44, dadurch gekennzeichnet, dass das jeweils gleiche Testsignal wiederholt mit unterschiedlicher Lautstärke erzeugt wird, wobei nach Maßgabe der Bewertung durch den Hörgeräteträger die Kennlinien der spektral selektiven Dynamikkompression eingestellt werden.

46. Verfahren nach einem der Ansprüche 43 bis 45, dadurch gekennzeichnet, dass das Sprachelement ein Vokal ist.

47. Verfahren nach einem der Ansprüche 43 bis 46, dadurch gekennzeichnet, dass in einem weiteren Verfahrensschritt ein Testsignal erzeugt wird, das einem natürlichen Frikativ entspricht.

48. Verfahren nach einem der Ansprüche 43 bis 47, dadurch gekennzeichnet, dass in einem weiteren Verfahrensschritt ein Testsignal erzeugt wird, das einem natürlichen Plosiv entspricht.

49. Verfahren nach einem der Ansprüche 43 bis 48, dadurch gekennzeichnet, dass in einem weiteren Verfahrensschritt Testsignale erzeugt werden, die verschiedenen Vokalen mit hoch liegenden zweiten Formanten entsprechen, wobei der Hörgeräteträger die Unterscheidbarkeit der Testsignale bewertet.

50. Verfahren nach Anspruch 49, dadurch gekennzeichnet, dass nach Maßgabe der Bewertung die Dynamikkennlinien in den spektralen Bereichen der hoch liegenden zweiten Formanten eingestellt werden.

51. Verfahren nach einem der Ansprüche 43 bis 50, dadurch gekennzeichnet, dass gleichzeitig mit den Testsignalen Störlärmsignale erzeugt werden.

52. Vorrichtung (1) zum Verarbeiten von akustischen Sprachsignalen (3), mit einer elektronischen Verarbeitungseinrichtung, dadurch gekennzeichnet, dass die Verarbeitungseinrichtung zum lautklassenspezifischen Verarbeiten der Sprachsignale (3) eingerichtet ist und Mittel aufweist, mit denen eine zeitliche Verlängerung von schwach artikulierten Lauten durchführbar ist.

53. Vorrichtung (1) nach Anspruch 52, gekennzeichnet durch eine Filtereinrichtung (2), mittels der die Sprachsignale (3) in oberhalb einer oberen Grenzfrequenz liegende, hochfrequente Frequenzbänder (FB4, FB5) und in unterhalb der oberen Grenzfrequenz liegende Frequenzbänder (FB1, FB2, FB3) aufspaltbar sind.

54. Vorrichtung (1) nach Anspruch 53, dadurch gekennzeichnet, dass die Filtereinrichtung (2) mehrere Filtereinheiten aufweist, mit denen jeweils ein bestimmtes unter der oberen Grenzfrequenz liegendes Frequenzband (FB1, FB2, FB 3) individuell mit einer Pre-Emphasis modulierbar ist.

55. Vorrichtung (1) nach einem der Ansprüche 52 bis 54, gekennzeichnet durch einen Lautklassenselektor (9), mit dem einem Sprachsignal (3) eine bestimmte Lautklasse zugeordnet werden kann.

56. Vorrichtung (1) nach einem der Ansprüche 53 bis 55, gekennzeichnet durch wenigstens eine Frequenzbandverschiebeeinheit (7), mit der ein hochfrequentes Frequenzband (FB4, FB5) individuell zu niedrigeren Frequenzen unterhalb der oberen Grenzfrequenz verschiebbar ist.

57. Vorrichtung (1) nach einem der Ansprüche 53 bis 56, gekennzeichnet durch eine Kompressoreinheit (19), mittels der die Frequenzbänder individuell komprimierbar sind.

58. Vorrichtung (1) nach einem der Ansprüche 52 bis 57, gekennzeichnet durch einen Equalizer (23), mit dem eine Kompensation von ungewünschten spektralen Eigenschaften von an die Verarbeitungseinrichtung anschließbaren Sprachsignalausgabeeinheiten durchführbar ist.

59. Vorrichtung (1) nach Anspruch 57 oder 58, dadurch gekennzeichnet, dass die zeitliche Verlängerung der schwach artikulierten Laute, die Verschiebung der hochfrequenten Frequenzbänder (FB4, FB5) und/oder die individuelle Kompression der Frequenzbänder mittels des Lautklassenselektors (9) steuerbar sind.

60. Vorrichtung zum Verarbeiten von akustischen Sprachsignalen, mit einer elektronischen Verarbeitungseinrichtung,
d a d u r c h g e k e n n z e i c h n e t,
dass die Verarbeitungseinrichtung zum Ersetzen von schwach artikulierten Lauten durch ihnen entsprechende, synthetische Laute eingerichtet ist.

61. Sprachsynthesizer mit
- Mitteln zur Erzeugung von zwei oder mehr Formantwellenformen jeweils durch Modulation eines bei einer Formantfrequenz oszillierenden Quellensignals mit einer Umhüllenden-Funktion,
- Mitteln zur Addition der zwei oder mehr Formantwellenformen, und
- Mitteln zur Überlagerung und Verkettung der addierten Formantwellenformen gemäß einer Pitchintervalllänge zu einem Sprachsignal,
d a d u r c h g e k e n n z e i c h n e t,
dass der Sprachsynthesizer zur Durchführung des Verfahrens nach einem der Ansprüche 25 bis 42 eingerichtet ist.

62. Anordnung zur Kontrolle der Anpassung eines Hörgerätes (407) an ein Hördefizit eines Hörgeräteträgers (406), umfassend:
- einen Personalcomputer (401);
- ein mit dem Personalcomputer (401) verbundenes Audio-Interface (402);
- wenigstens ein mit dem Audio-Interface (402) verbundener Lautsprecher (405),

gekennzeichnet durch eine auf dem Personalcomputer (401) ablaufende Programmsteuerung mit Instruktionen zur Durchführung des Verfahrens nach einem der Ansprüche 43 bis 51.

63. Anordnung nach Anspruch 62, gekennzeichnet durch ein Linearisierungsfilter, mittels welchem der Frequenzgang des wenigstens einen Lautsprechers (405) bei der Ausgabe von Testsignalen linearisiert wird.

64. Anordnung nach Anspruch 63, gekennzeichnet durch ein mit dem Audio-Interface (402) verbundenes Mikrophon (409) zur Kalibrierung des Linearisierungsfilters.

65. Computerprogramm zur Kontrolle der Anpassung eines Hörgerätes,
dadurch gekennzeichnet,
dass es Instruktionen zur Durchführung des Verfahrens nach einem der Ansprüche 43 bis 51 umfasst.

66. Verwendung eines Sprachsyntheseverfahrens nach einem der Ansprüche 25 bis 42 zur Erzeugung von Testsignalen in einem Verfahren nach einem der Ansprüche 43 bis 51.

- Zusammenfassung -

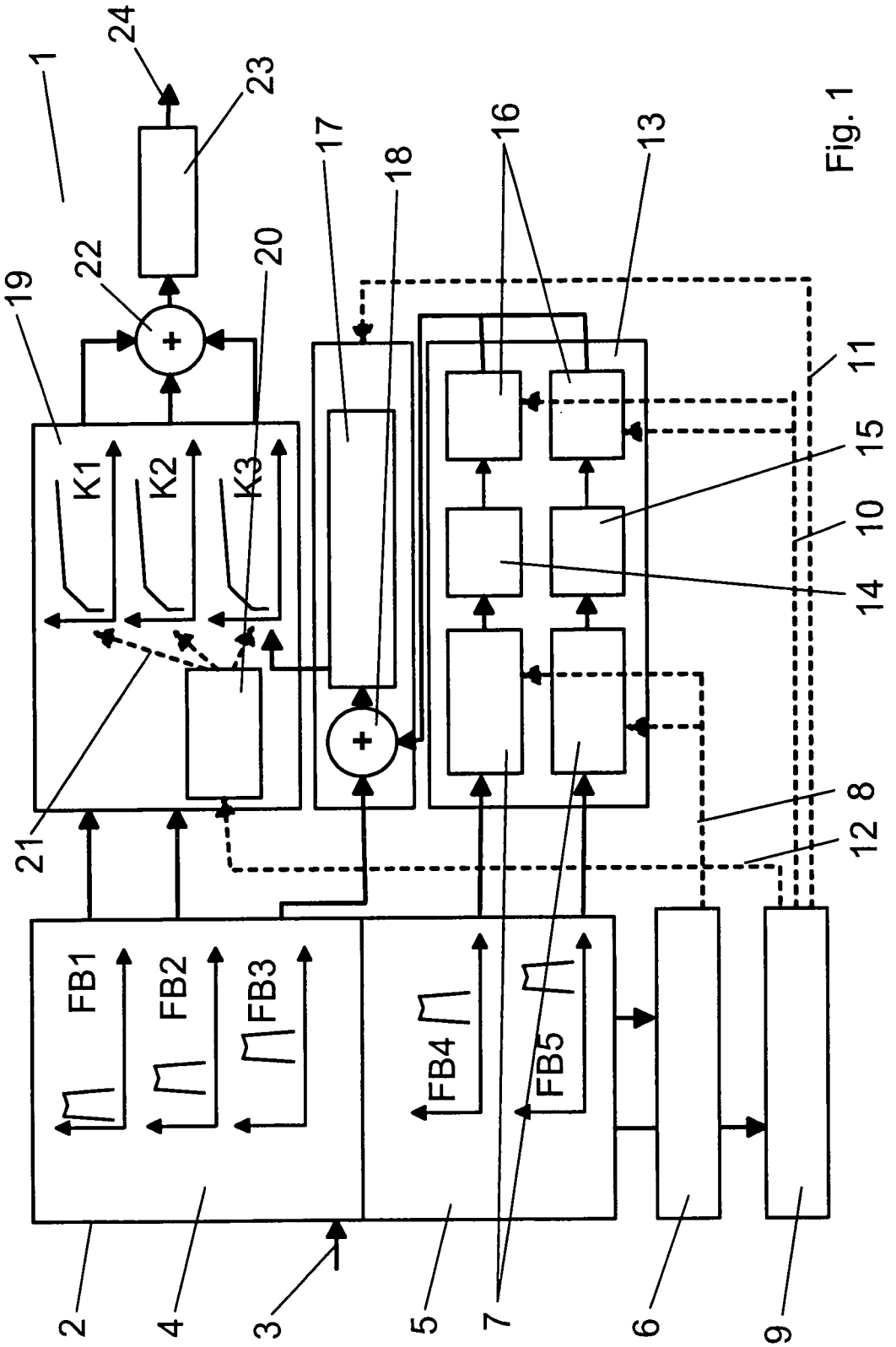

1
24
23
17
18
16
13
22
20
19
11
K1
K2
K3
15
10
14
21
8
12
FB1
FB2
FB3
FB4
FB5
2
4
3
5
7
6
9

Fig. 1

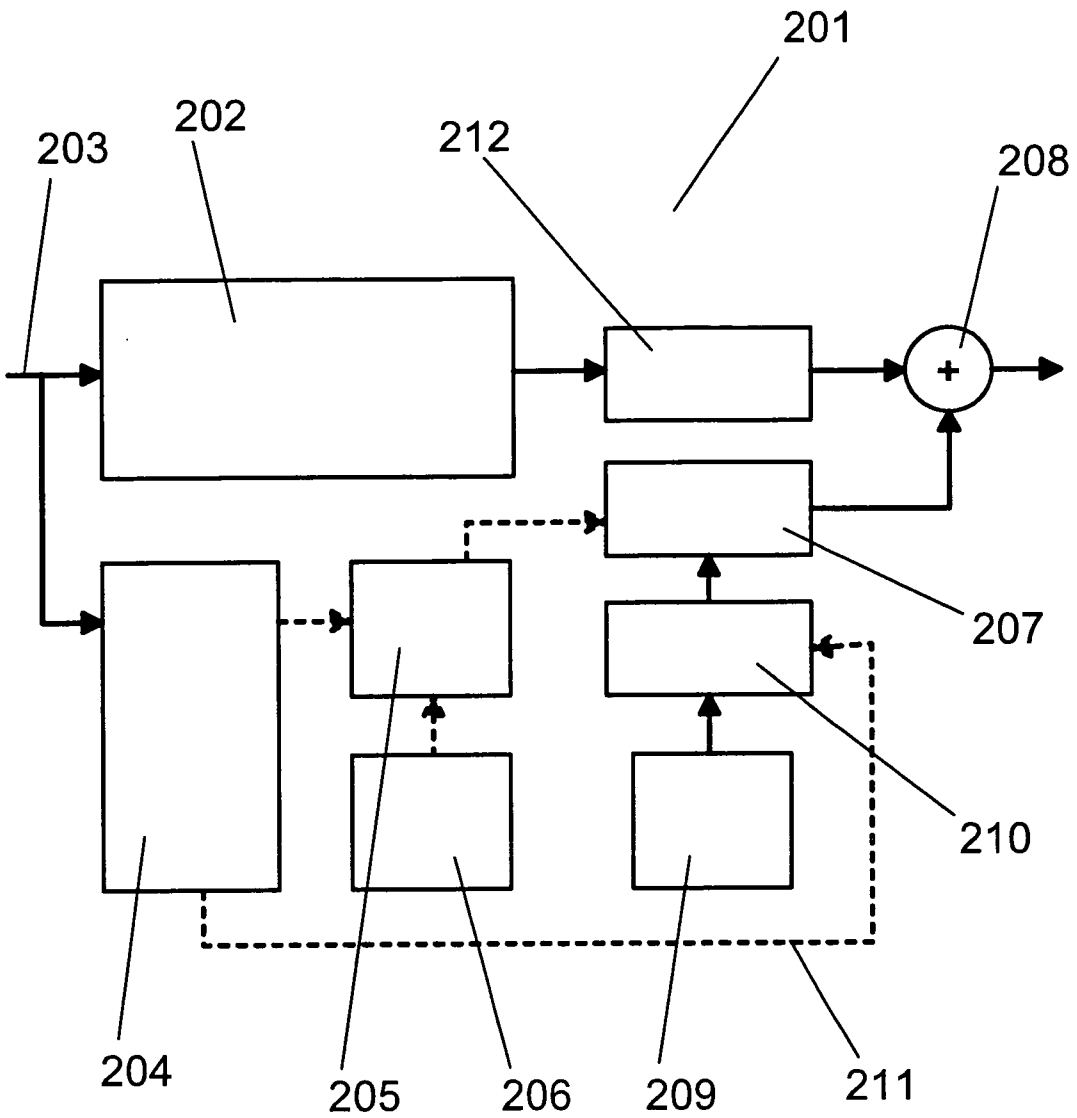


Fig. 2

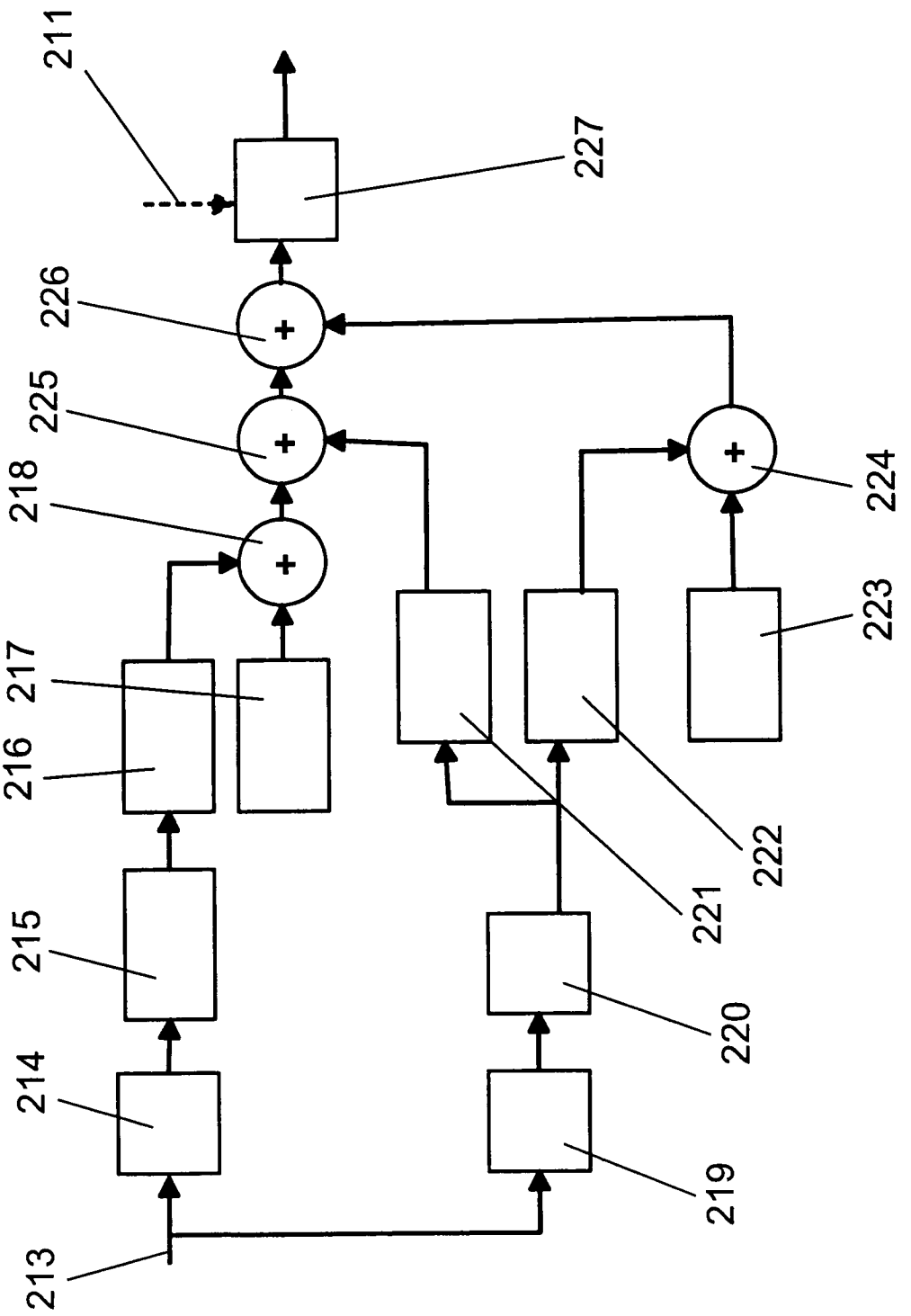


Fig. 3

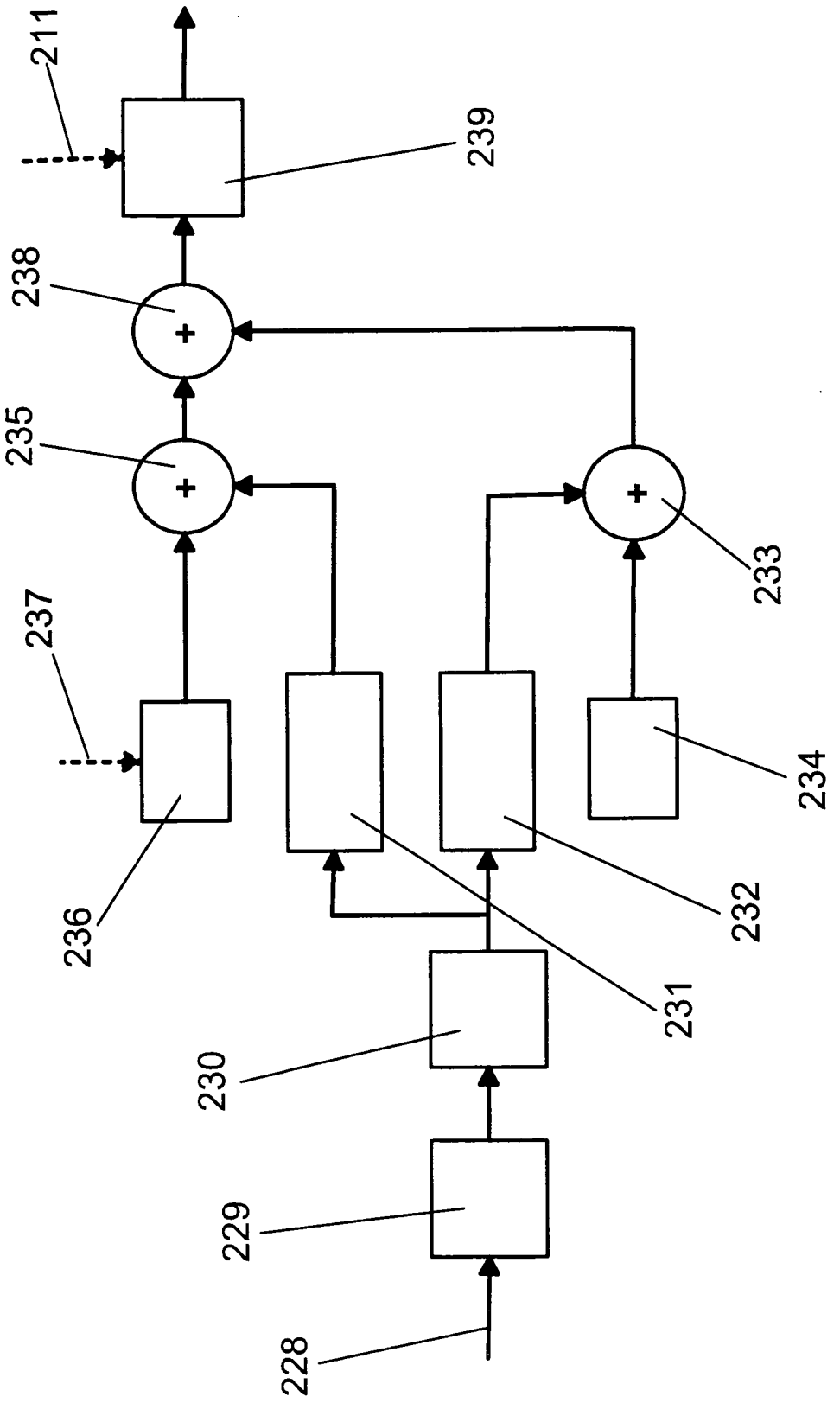


Fig. 4

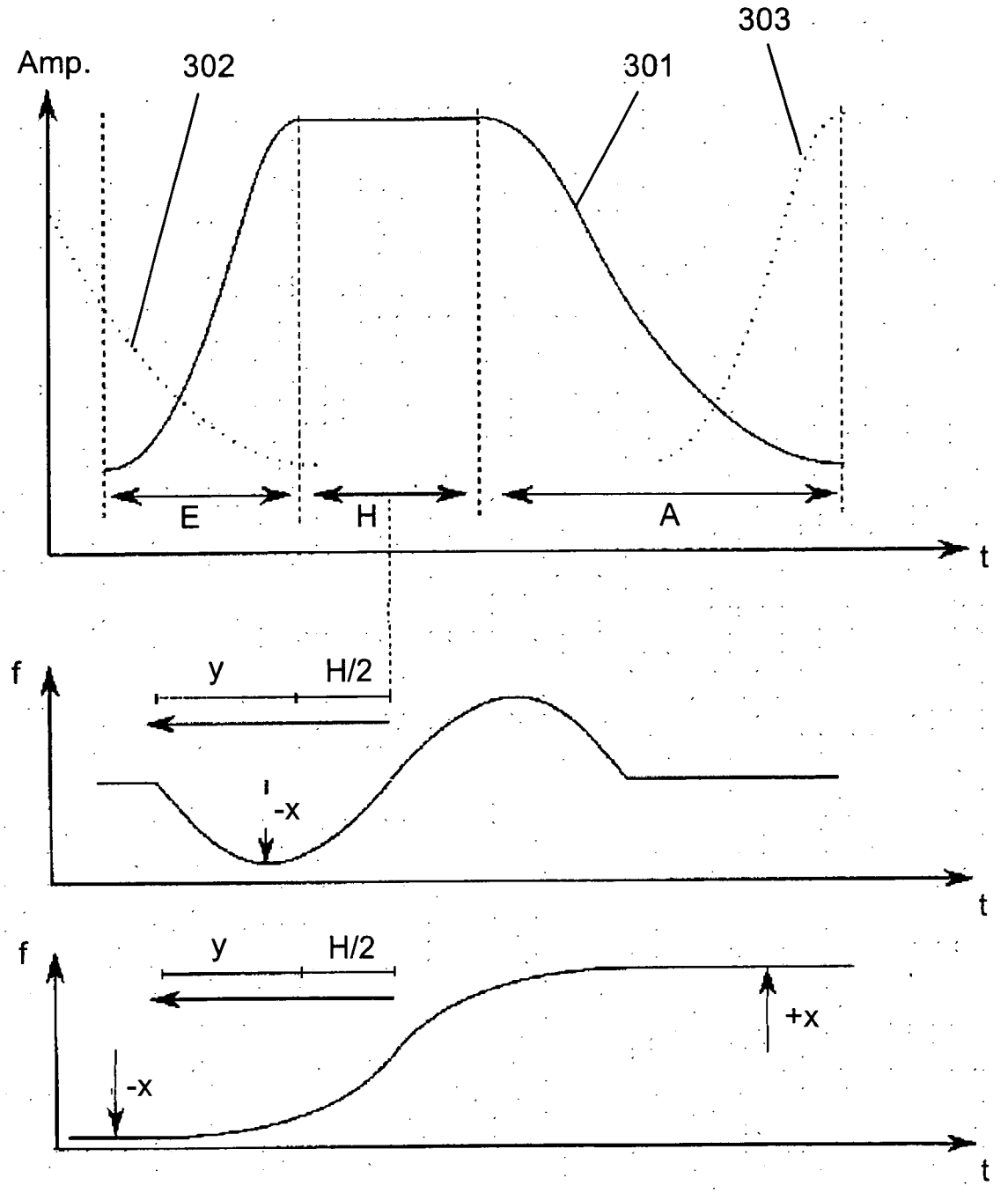
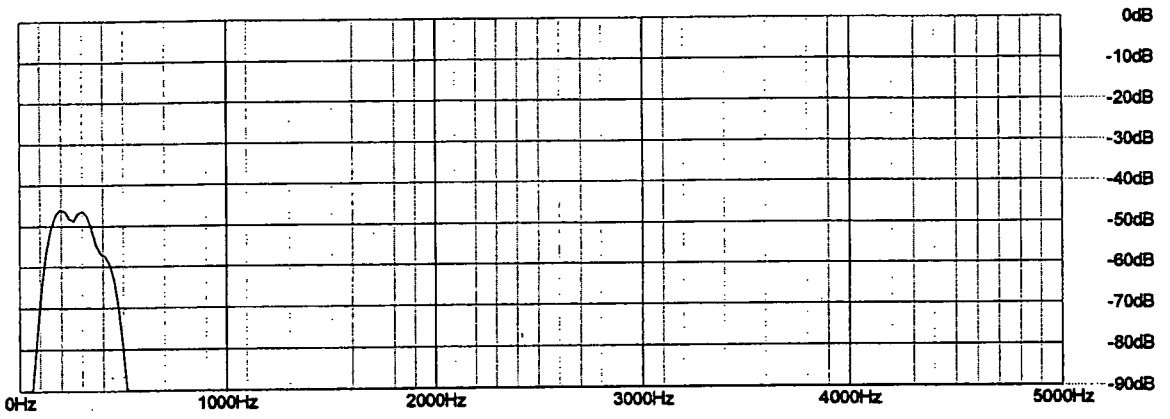
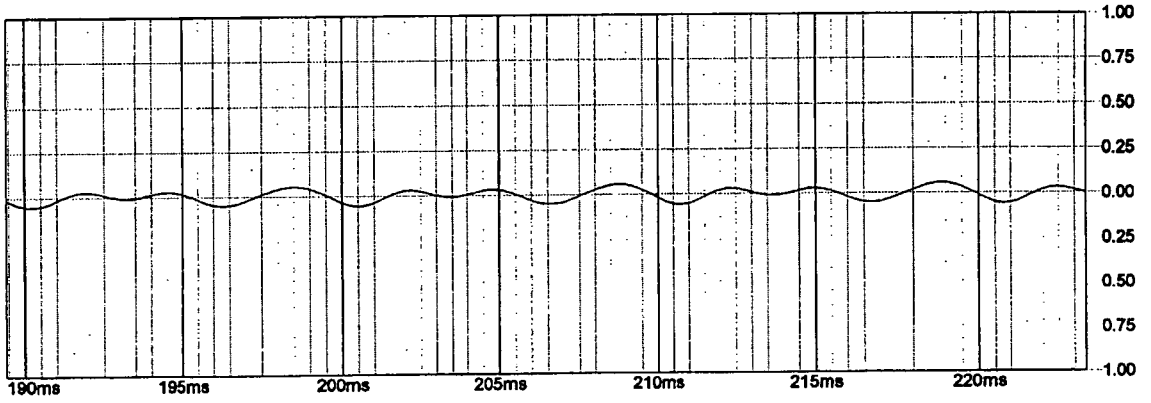


Fig. 5

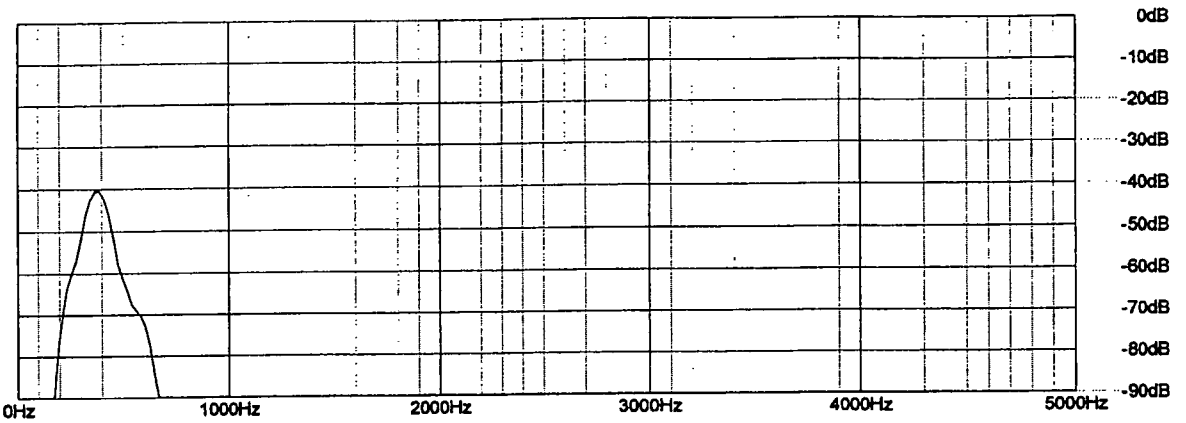

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

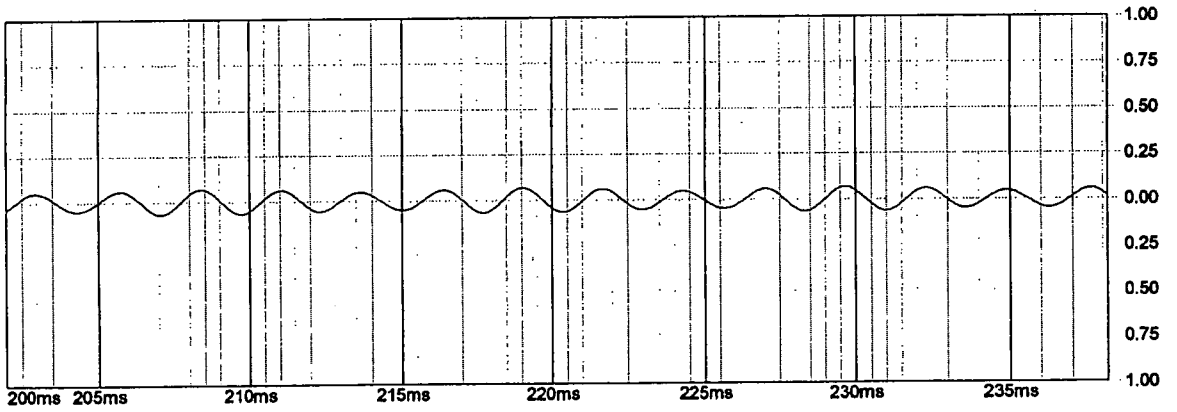

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
190ms
195ms
200ms
205ms
210ms
215ms
220ms

b)

Fig. 6

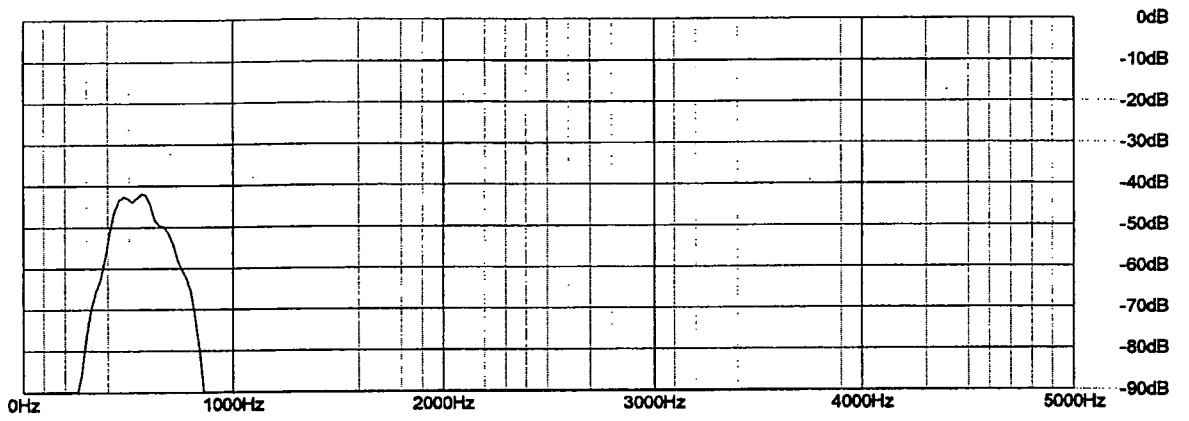

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

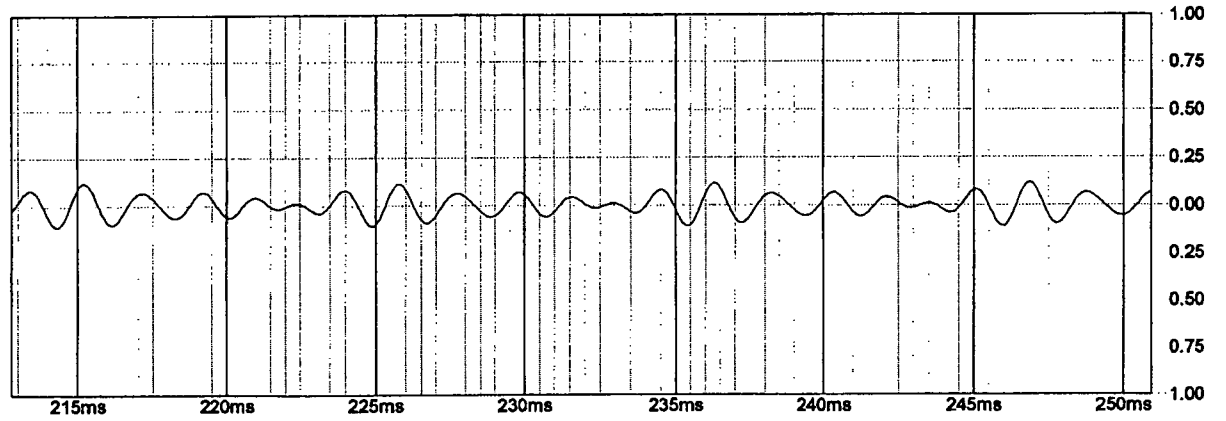

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
200ms
205ms
210ms
215ms
220ms
225ms
230ms
235ms

b)

Fig. 7

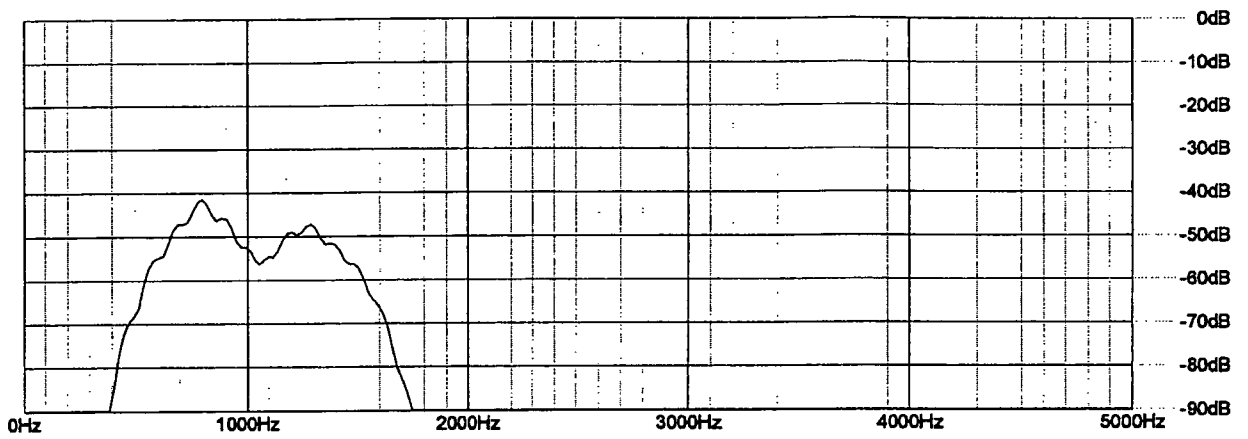

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

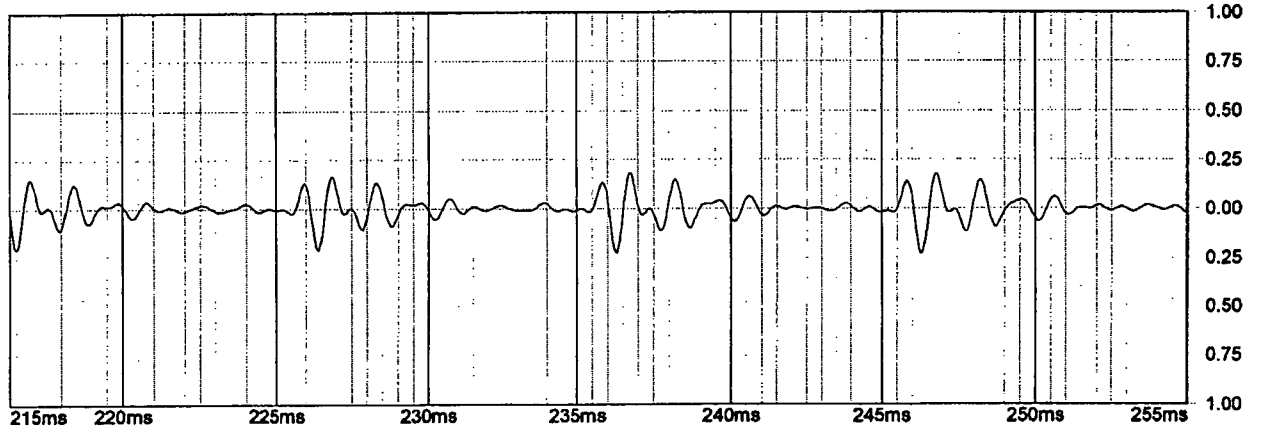

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
215ms
220ms
225ms
230ms
235ms
240ms
245ms
250ms

b)

Fig. 8

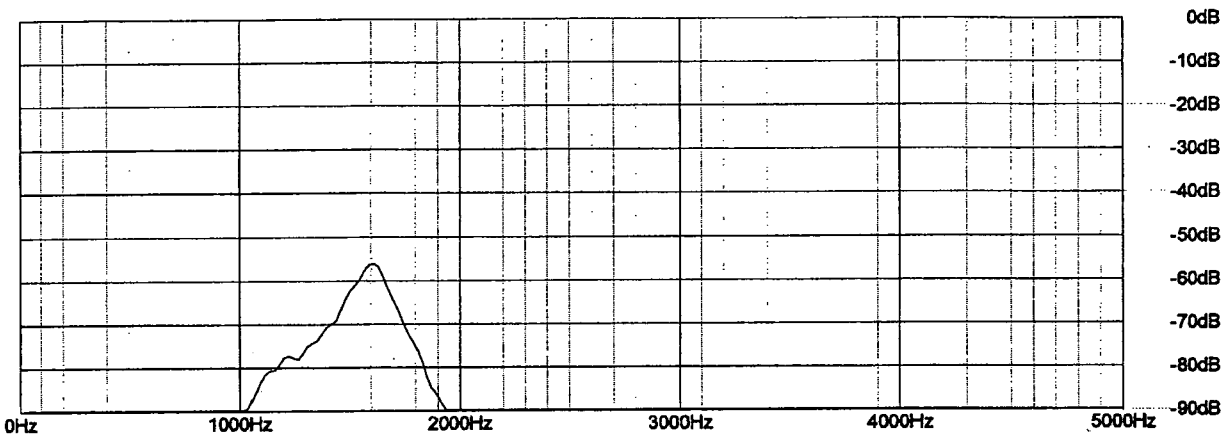

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

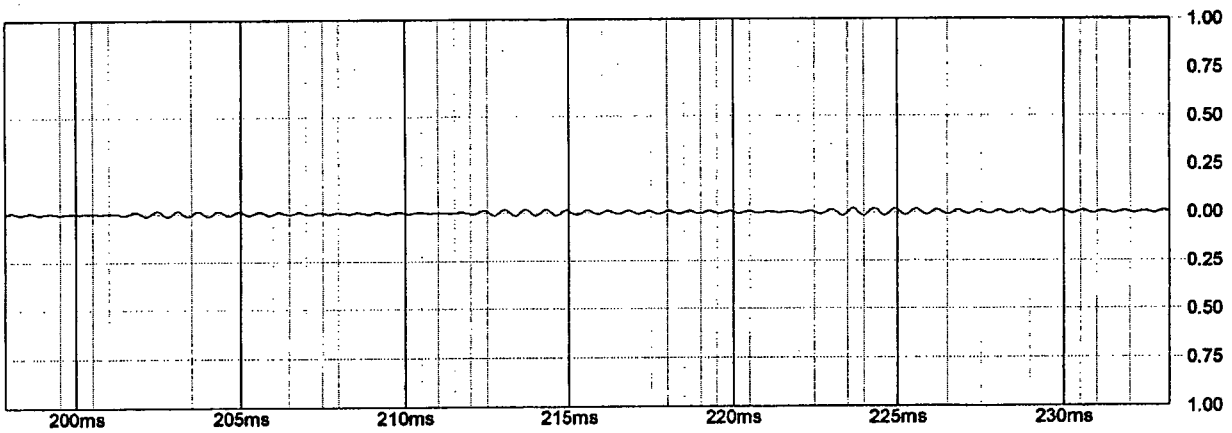

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
215ms
220ms
225ms
230ms
235ms
240ms
245ms
250ms
255ms

b)

Fig. 9

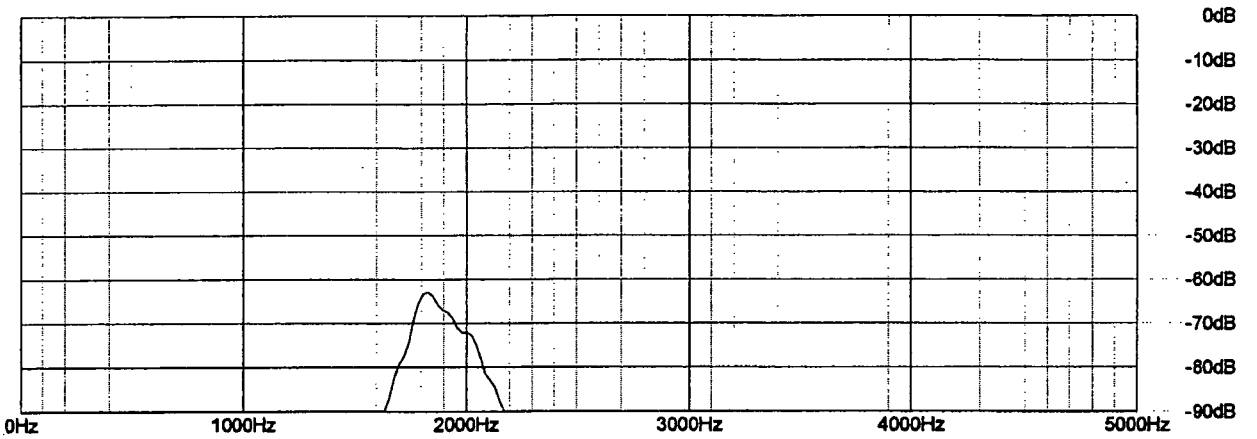

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

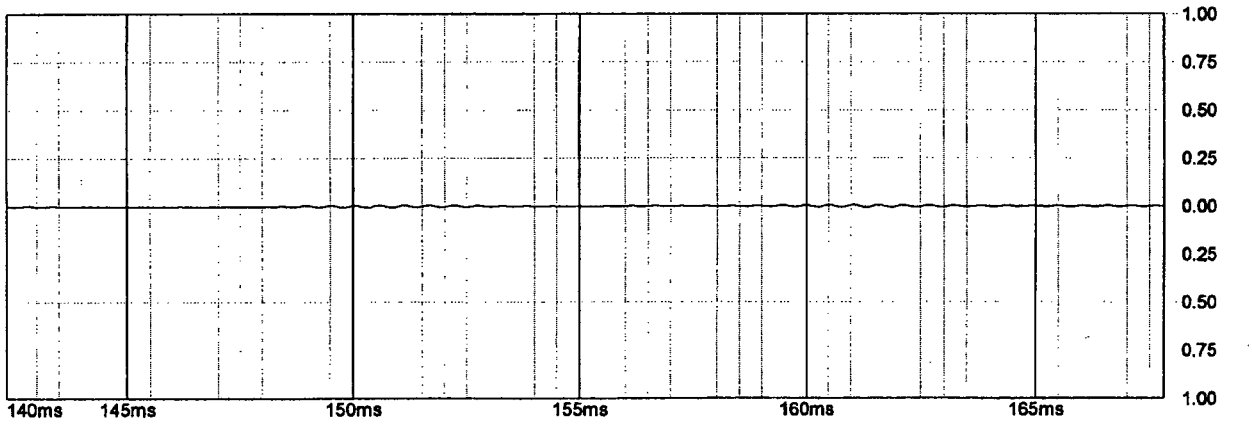

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
200ms
205ms
210ms
215ms
220ms
225ms
230ms

b)

Fig. 10

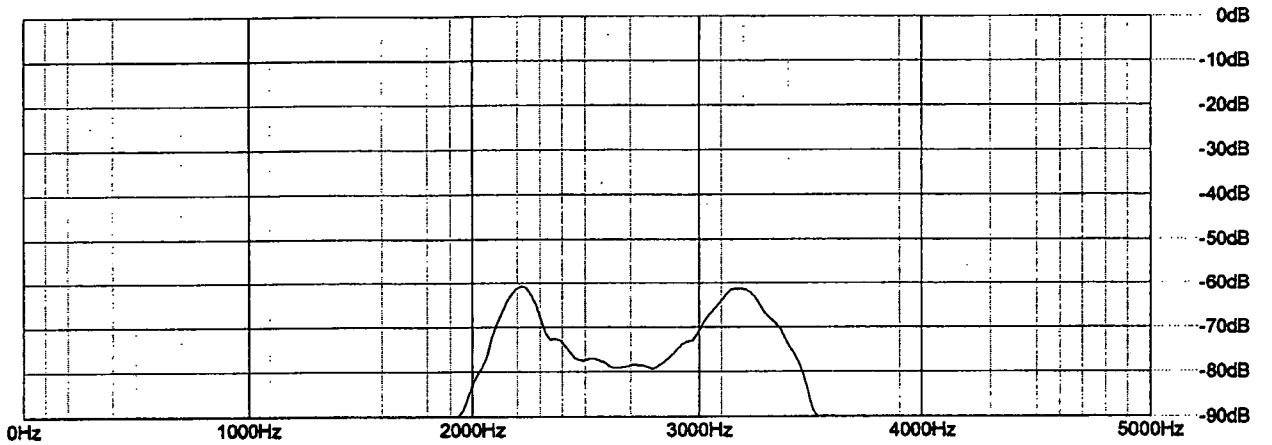

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

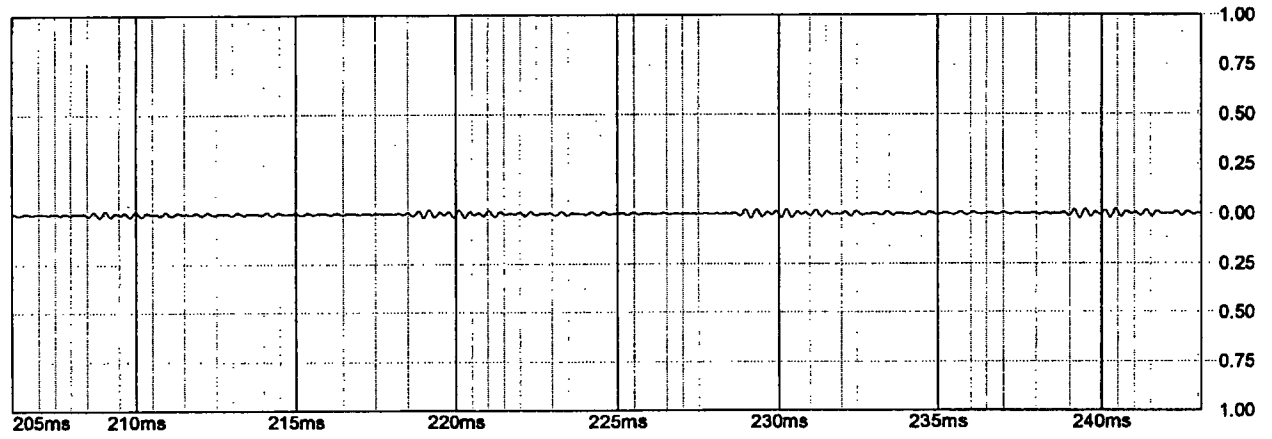

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
140ms
145ms
150ms
155ms
160ms
165ms

b)

Fig. 11

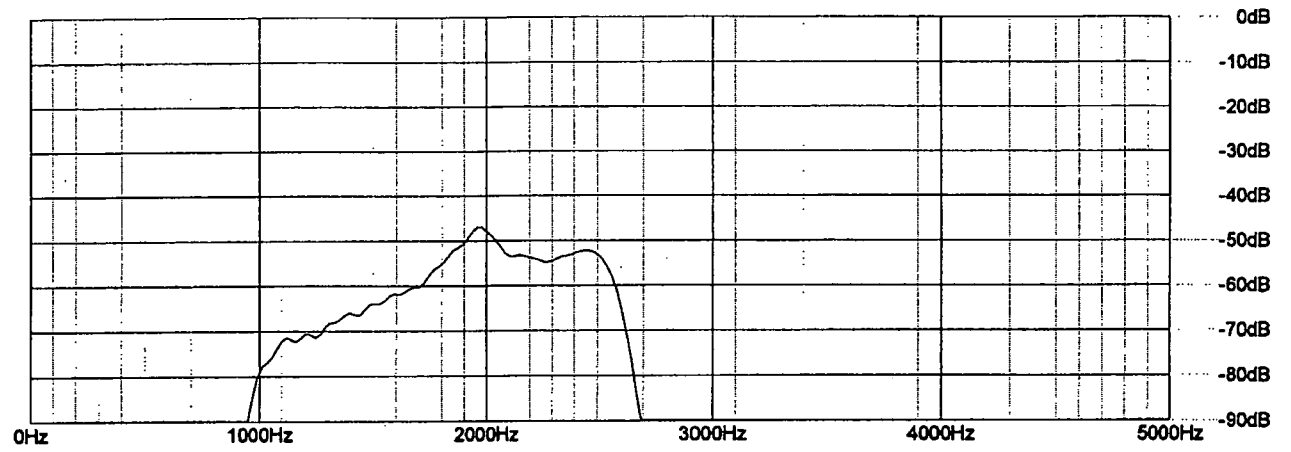

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

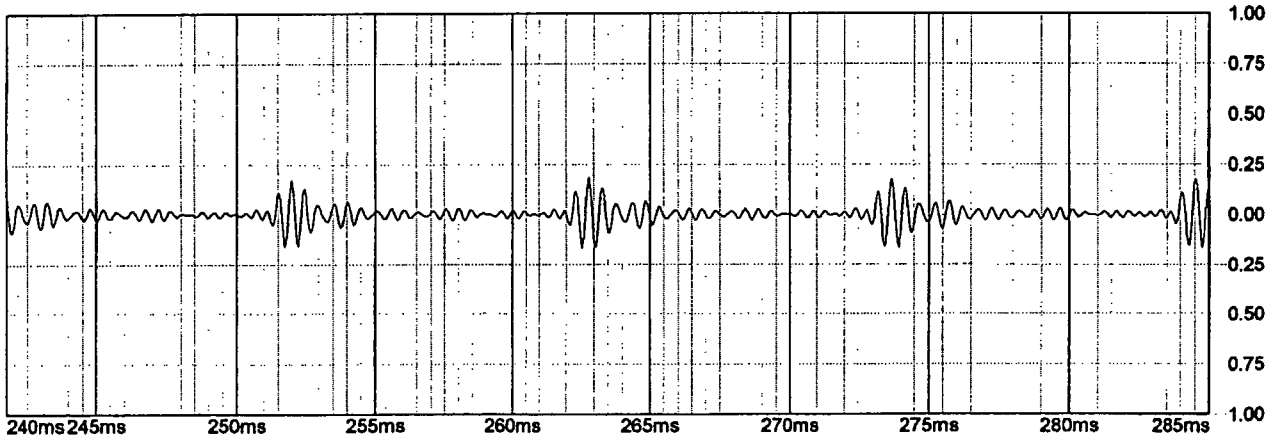

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
205ms
210ms
215ms
220ms
225ms
230ms
235ms
240ms

b)

Fig. 12

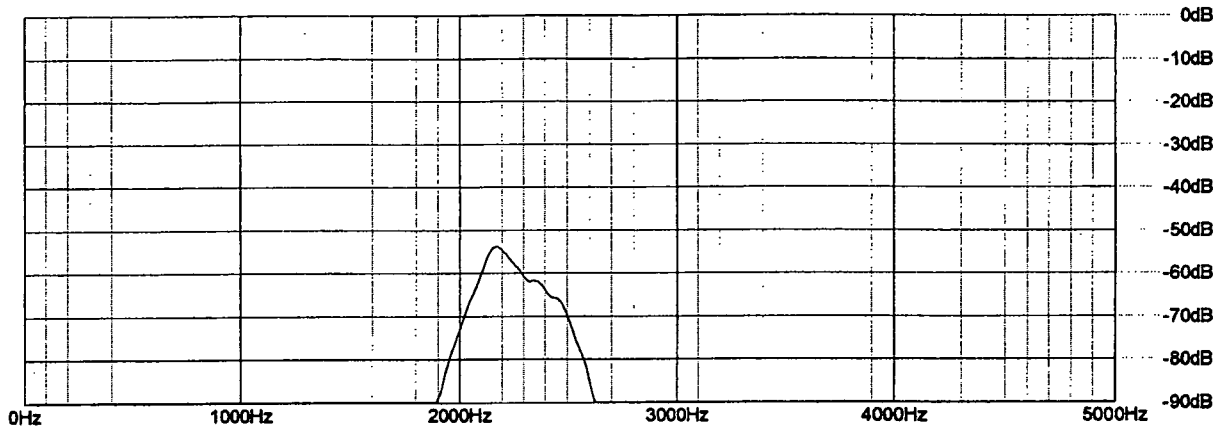

0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)

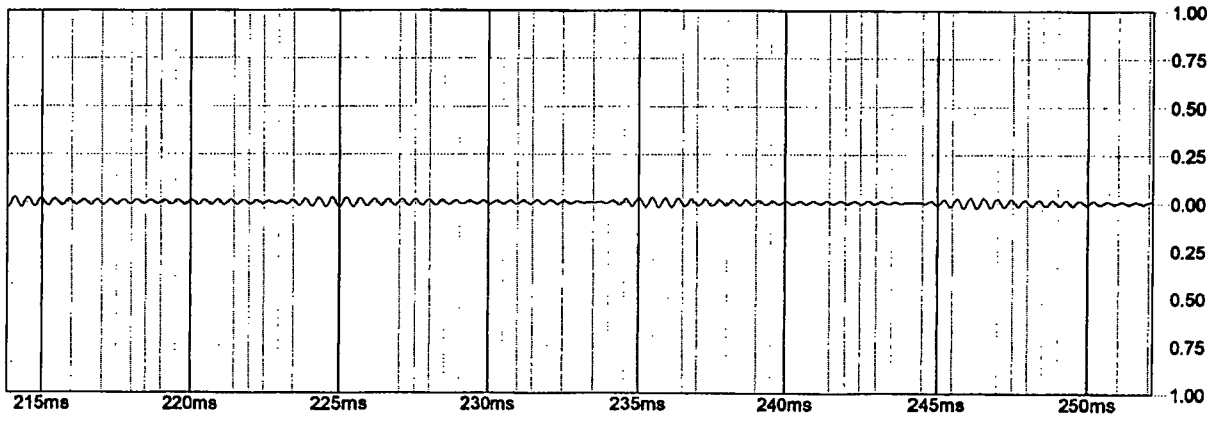

1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
240ms
245ms
250ms
255ms
260ms
265ms
270ms
275ms
280ms
285ms

b)

Fig. 13


0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

a)


1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
215ms
220ms
225ms
230ms
235ms
240ms
245ms
250ms

b)

Fig. 14

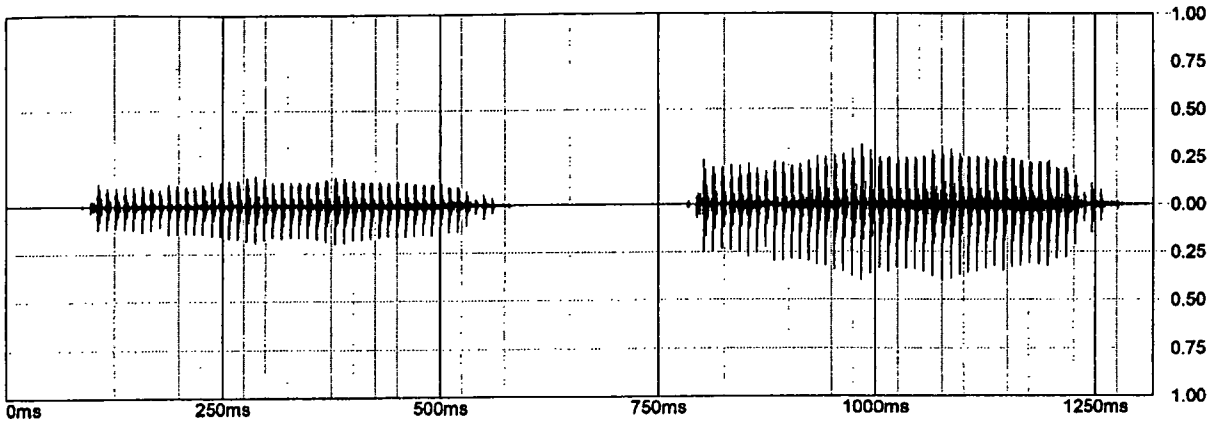


Fig. 15

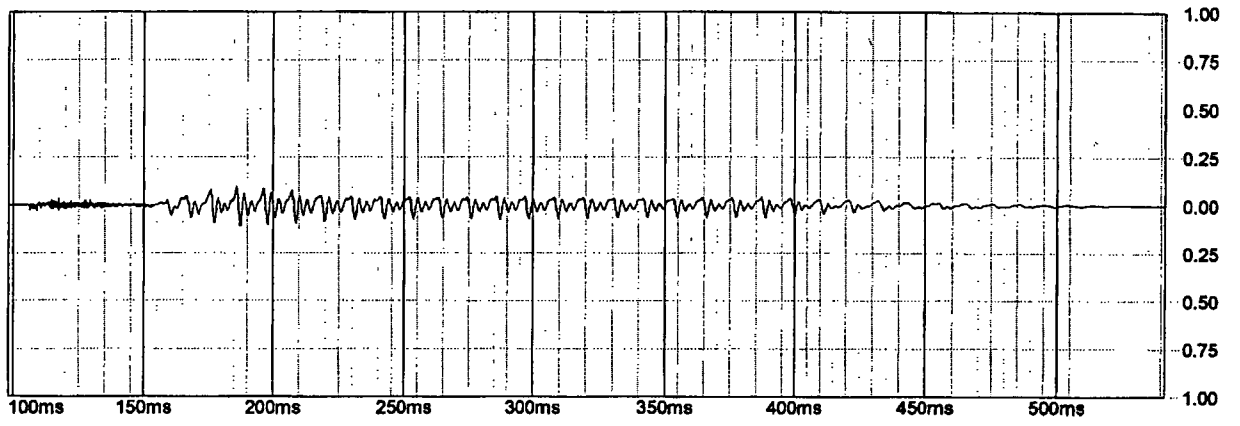


Fig. 17

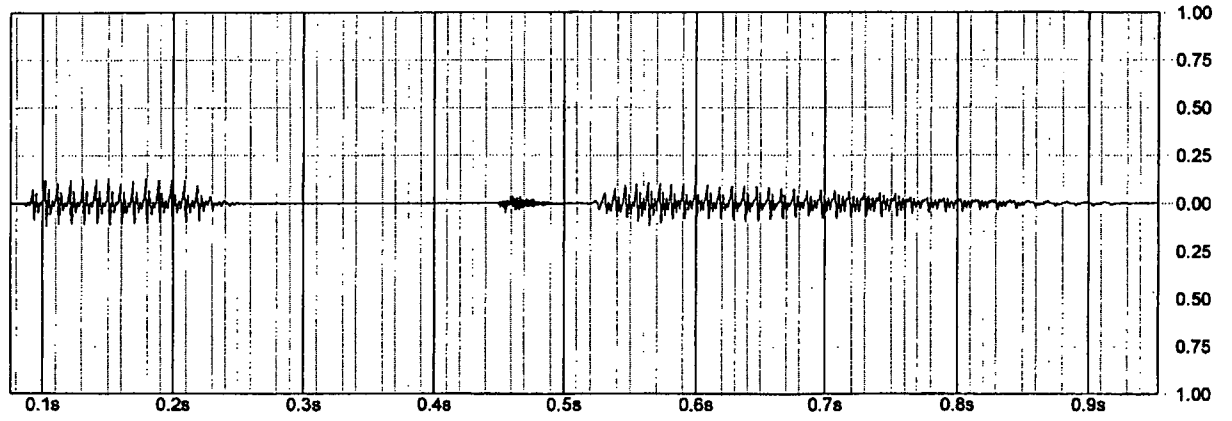
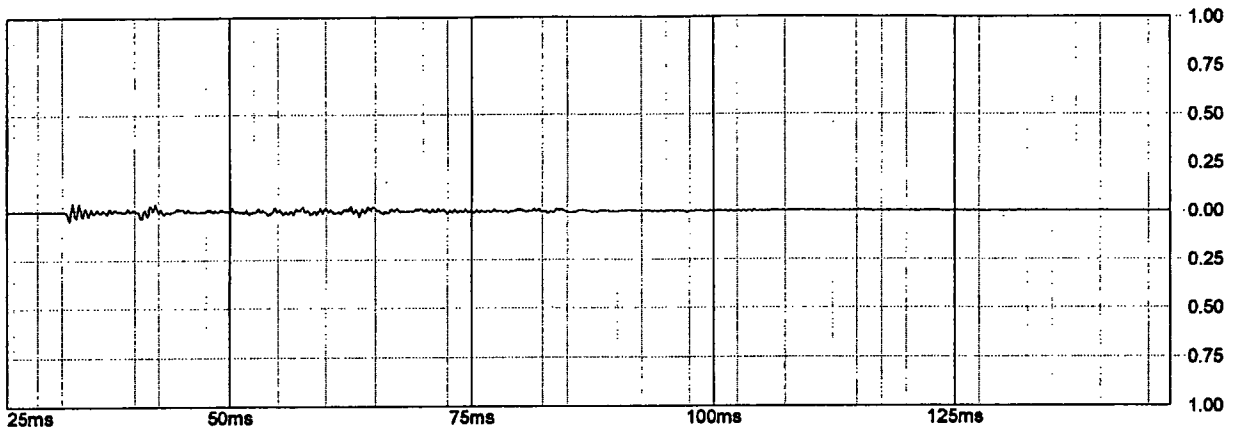
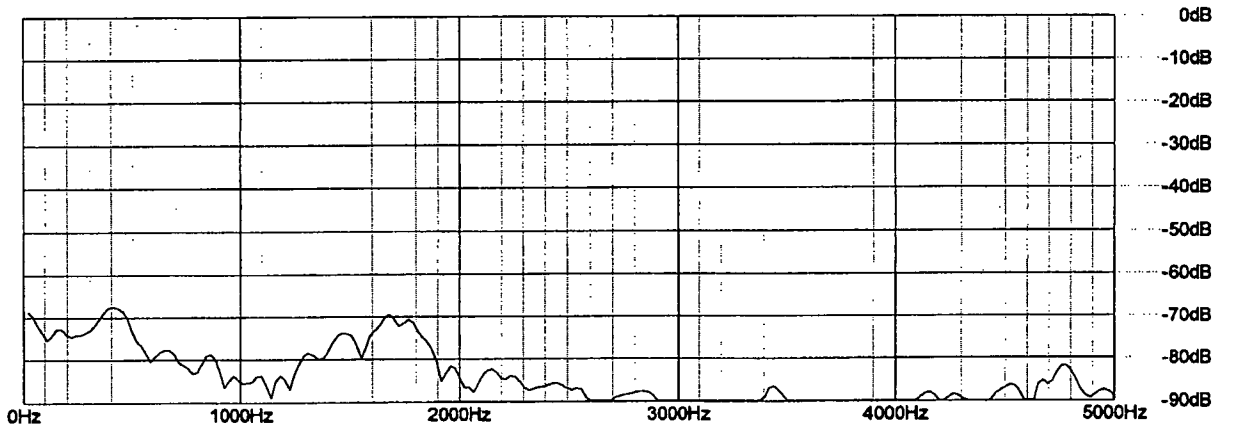


Fig. 18


1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
25ms
50ms
75ms
100ms
125ms

a)


0dB
-10dB
-20dB
-30dB
-40dB
-50dB
-60dB
-70dB
-80dB
-90dB
0Hz
1000Hz
2000Hz
3000Hz
4000Hz
5000Hz

b)

Fig. 16

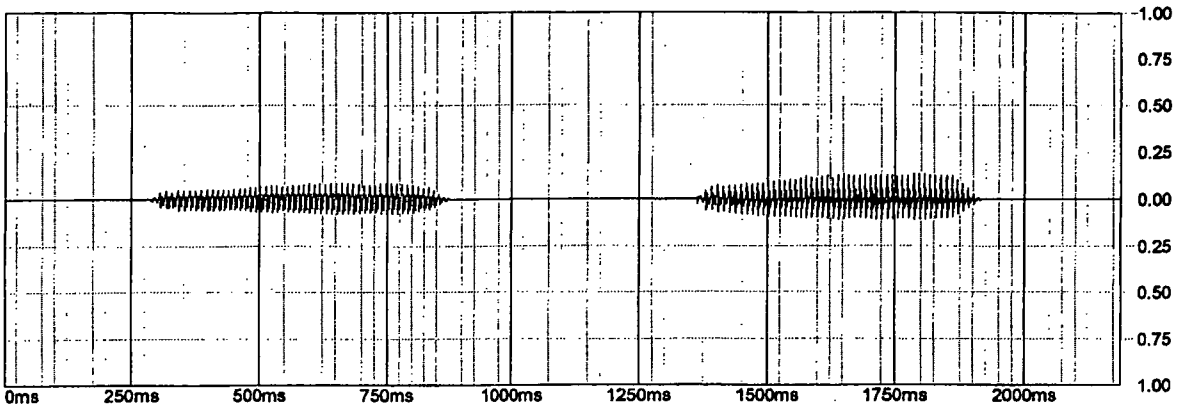
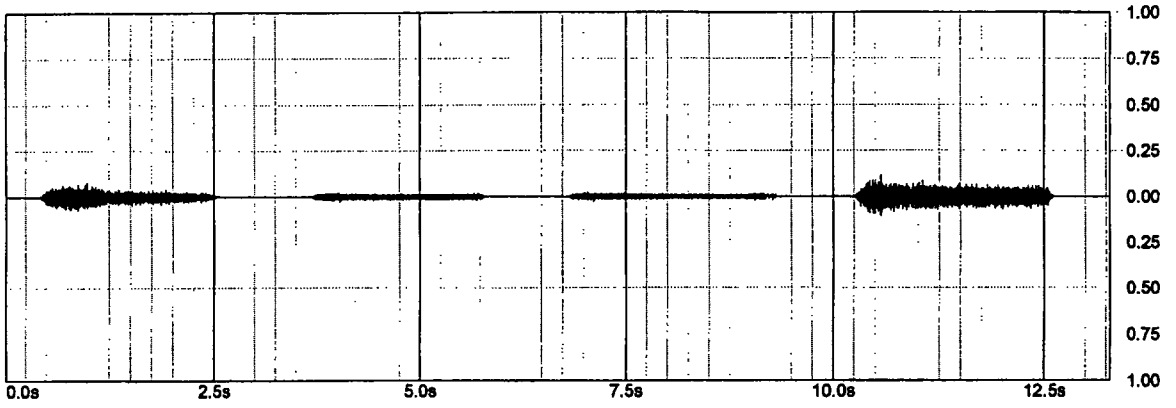
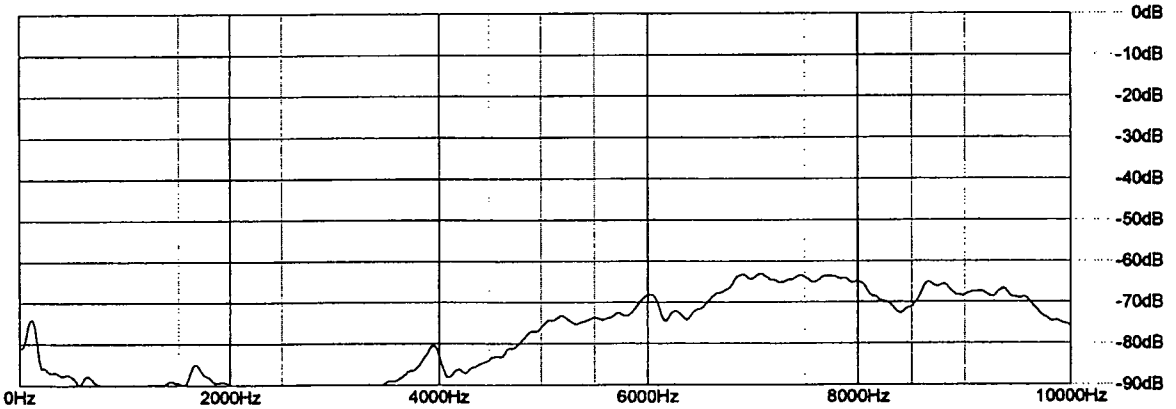


Fig. 19



a)



b)

Fig. 20

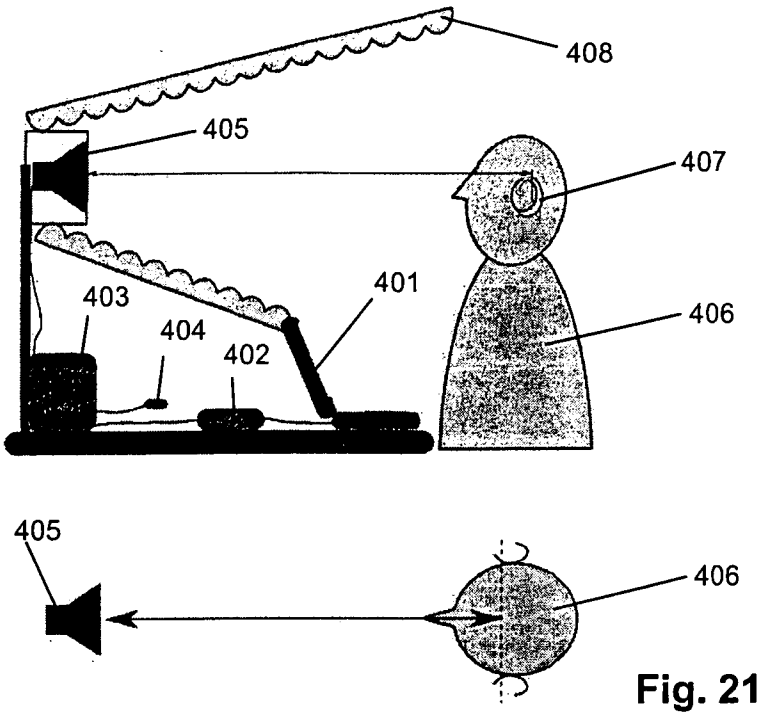


Fig. 21

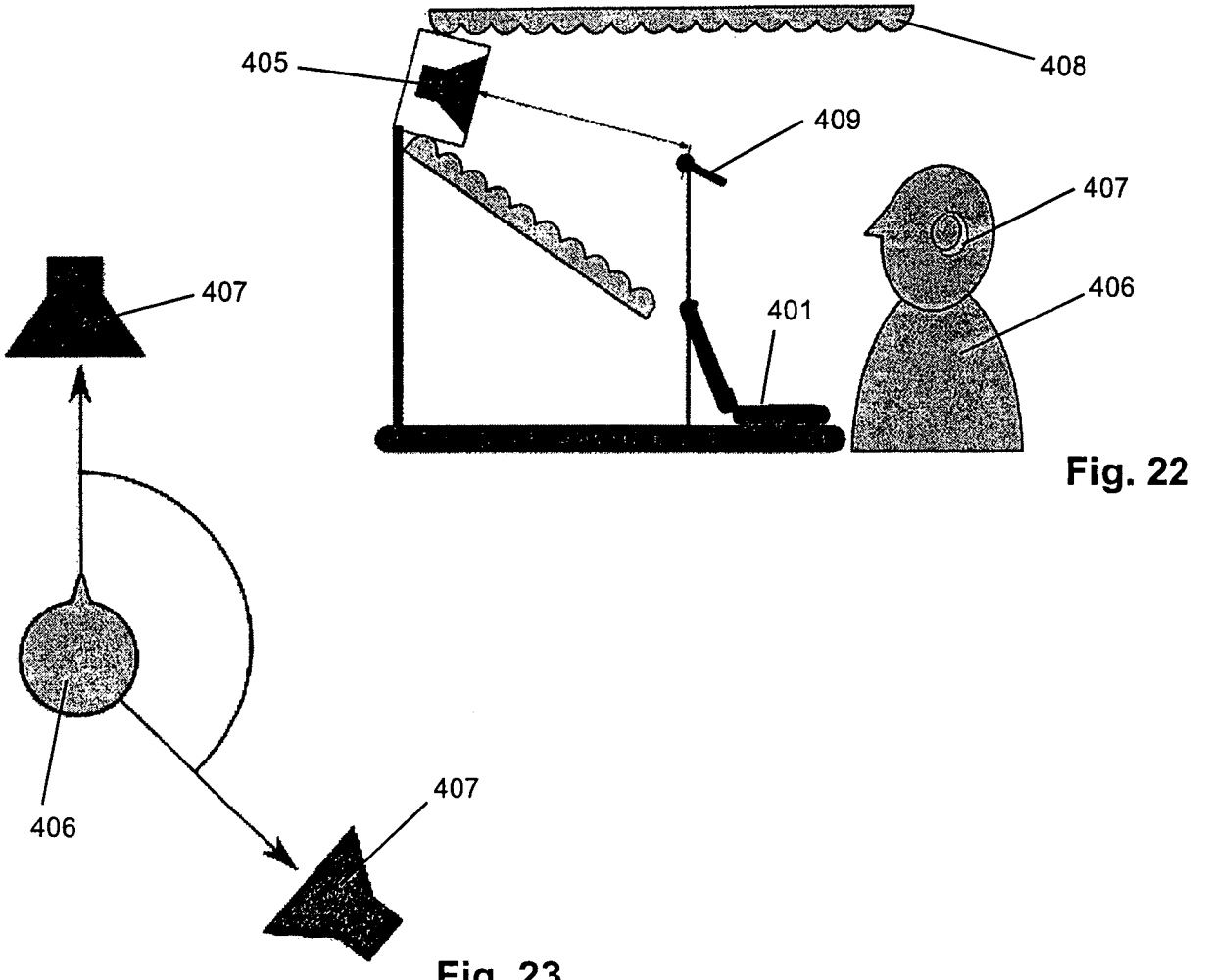


Fig. 22

Fig. 23