(19) **Bundesrepublik Deutschland**
Deutsches Patent- und Markenamt

(10) **DE 602 02 161 T2** 2005.12.15

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 1 239 463 B1**
(21) Deutsches Aktenzeichen: **602 02 161.8**
(96) Europäisches Aktenzeichen: **02 005 150.4**
(96) Europäischer Anmeldetag: **07.03.2002**
(97) Erstveröffentlichung durch das EPA: **11.09.2002**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **08.12.2004**
(47) Veröffentlichungstag im Patentblatt: **15.12.2005**

(51) Int Cl.[7]: **G10L 13/02**
G10L 19/02

(30) Unionspriorität:
2001067257 09.03.2001 JP

(84) Benannte Vertragsstaaten:
DE, GB

(73) Patentinhaber:
YAHAMA CORPORATION, Hamamatsu, Shizuoka, JP

(74) Vertreter:
WAGNER & GEYER Partnerschaft Patent- und Rechtsanwälte, 80538 München

(72) Erfinder:
Yoshioka, Yasuo, Hamamatsu-shi, Shizuoka, JP;
Sanjaume, Bonada, Jordi, 08003 Barcelona, ES

(54) Bezeichnung: **Verfahren, Vorrichtung und Programm zur Analyse und Synthese von Sprache**

Beschreibung

Hintergrund der Erfindung

A) Gebiet der Erfindung

[0001] Die vorliegende Erfindung betrifft eine Stimmensynthesevorrichtung und im speziellen eine Stimmensynthesevorrichtung zum Synthetisieren bzw. zum künstlichen Herstellen von Stimmen eines Liedes, das von einem Sänger gesungen wird.

B) Beschreibung des verwandten Standes der Technik

[0002] Menschliche Stimmen bestehen aus Phonemen, von denen jedes aus einer Vielzahl von Formanten gebildet werden. Beim Synthetisieren von Stimmen eines Liedes, das durch einen Sänger gesungen wird, werden zuerst alle Formanten, die all diejenigen Phoneme bilden, die von einem Sänger erzeugt werden können, generiert und synthetisiert, um jedes Phonem zu formen. Als Nächstes wird eine Vielzahl der erzeugten Phoneme sequenziell bzw. in Folge gekoppelt, und Tonhöhen werden in Einklang mit der Melodie gesteuert, um dadurch Stimmen eines von einem Sänger gesungenen Liedes zu synthetisieren. Dieses Verfahren ist nicht nur auf menschliche Stimmen anwendbar, sondern auch auf musikalische Klänge, die durch ein musikalisches Instrument, wie beispielsweise ein Blasinstrument, erzeugt werden.

[0003] Eine Stimmensynthesevorrichtung, welche dieses Verfahren verwendet, ist bereits bekannt. Beispielsweise offenbart das Japanische Patent Nr. 2504172 eine Formantenklang-erzeugende Vorrichtung, welche einen Formantenklang erzeugen kann, der sogar eine hohe Tonlage aufweisen kann, ohne unnötige Spektren zu erzeugen.

[0004] Die oben beschriebene formantenklang-erzeugende Vorrichtung und die herkömmliche Stimmen synthetisierende bzw. Stimmen erzeugende Vorrichtung können keine individuellen Charakteristiken wie beispielsweise die Stimmenqualität, Eigentümlichkeit und ähnliches jeder Person reproduzieren, falls nur die Tonhöhe verändert wird, obwohl sie pseudonymartig Stimmen eines Liedes synthetisieren können, welches von einer allgemeinen Person gesungen wird.

[0005] Ein weiteres Beispiel einer bekannten Stimmensynthesevorrichtung wird in P. Cano et al.; "Voice Morphing for impersonating in Karaoke Applications", Proceedings of the International Computer Music Conference 2000, Berlin, Deutschland (2000), S. 1–4, offenbart.

Zusammenfassung der Erfindung

[0006] Es ist eine Aufgabe der vorliegenden Erfindung, so wie sie in den beigefügten Ansprüchen beansprucht wird, eine Stimmensynthesevorrichtung bereitzustellen, welche in der Lage ist, Stimmen eines von einem Sänger gesungenen Liedes zu synthetisieren und individuelle Charaktereigenschaften wie beispielsweise die Klangqualität, Eigentümlichkeit usw. jedes Sängers zu reproduzieren.

[0007] Es ist eine weitere Aufgabe der vorliegenden Erfindung, eine Stimmensynthesevorrichtung bereitzustellen, welche in der Lage ist, realistischere Stimmen eines von einem Sänger gesungenen Liedes zu synthetisieren und das Lied in einem natürlichen Zustand bzw. auf natürliche Art zu singen.

[0008] Gemäß eines Aspekts der vorliegenden Erfindung wird eine Stimmenanalysenvorrichtung bereitgestellt, welche folgendes umfasst: erste Analysemittel zum Analysieren bzw. Zerlegen einer Stimme in harmonische Komponenten und nicht-harmonische Komponenten und zweite Analysemittel zum Analysieren bzw. Zerlegen einer Größenspektrumumhüllenden der harmonischen Komponente in eine Größenspektrumumhüllende einer Stimmbandschwingungswellenform, Resonanzen sowie eine Spektrumumhüllende einer Differenz zwischen der Größenspektrumumhüllenden der harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen; und Mittel zum Speichern der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen sowie der Spektrumumhüllenden der Differenz.

[0009] Gemäß eines anderen Aspekts der Erfindung ist eine Stimmensynthesevorrichtung vorgesehen, welche umfasst: Mittel zum Speichern einer Größenspektrumumhüllenden einer Stimmbandschwingungswellenform, Resonanzen und einer Spektrumumhüllenden einer Differenz zwischen einer Größenspektrumumhüllen-

den einer harmonischen Komponente aus einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen, jeweils analysiert bzw. zerlegt aus den harmonischen Komponenten, die aus einer Stimme analysiert bzw. zerlegt worden sind, und nicht-harmonischen Komponenten, welche aus der Stimme analysiert bzw. zerlegt worden sind; Mittel zur Eingabe von Information über eine zu synthetisierende Stimme; Mittel zum Erzeugen einer flachen Größenspektrumumhüllenden; und Mittel zum Hinzufügen der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden für die Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz, die jeweils aus den Mitteln zum Speichern ausgelesen worden sind, zu der flachen Größenspektrumumhüllenden in Übereinstimmung mit der eingegebenen Information.

[0010] Gemäß eines weiteren Aspekts der Erfindung ist eine Stimmensynthesevorrichtung vorgesehen, welche umfasst: erste Analysemittel zum Analysieren bzw. Zerlegen einer Stimme in harmonische Komponenten und nicht-harmonische Komponenten; zweite Analysemittel zum Analysieren bzw. Zerlegen einer Größenspektrumumhüllenden der harmonischen Komponenten in eine Größenspektrumumhüllende einer Stimmbandschwingungswellenform, Resonanzen sowie eine Spektrumumhüllende einer Differenz zwischen der Größenspektrumumhüllenden der harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen, Mittel zum Speichern der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen sowie der Spektrumumhüllenden der Differenz; Mittel zur Eingabe von Information über eine zu synthetisierende Stimme; Mittel zum Erzeugen einer flachen Größenspektrumumhüllenden; und Mittel zum Hinzufügen der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden für die Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz, die jeweils aus den Mitteln zum Speichern ausgelesen wurden, zu der flachen Größenspektrumumhüllenden in Übereinstimmung mit der eingegebenen Information.

[0011] Wie oben ist es möglich, eine Stimmensynthesevorrichtung vorzusehen, welche in der Lage ist, menschliche Musikklänge zu synthetisieren und individuelle Charaktereigenschaften wie beispielsweise die Stimmqualität, Eigentümlichkeit usw. jeder Person wiederzugeben bzw. zu reproduzieren.

[0012] Es ist auch möglich, eine Stimmensynthesevorrichtung vorzusehen, welche in der Lage ist, realistische Stimmen eines von einem Sänger gesungenen Liedes zu synthetisieren und ein Lied auf natürliche Art zu singen.

Kurze Beschreibung der Zeichnungen

[0013] Fig. 1 ist ein Diagramm, das eine Stimmenanalyse nach einer Ausführungsform der Erfindung darstellt.

[0014] Fig. 2 ist ein Graph, der eine Spektrumumhüllende harmonischer Komponenten zeigt;

[0015] Fig. 3 ist eine Zeichnung, die eine Größenspektrumumhüllende inharmonischer Komponenten zeigt;

[0016] Fig. 4 ist ein Graph, der Spektrumumhüllende einer Stimmbandschwingungswellenform zeigt.

[0017] Fig. 5 ist ein Graph, der eine Änderung in der Anregungskurve zeigt.

[0018] Fig. 6 ist ein Graph, der die durch VocalTractResonance gebildete Spektrumumhüllenden zeigt.

[0019] Fig. 7 ist ein Graph, der eine Spektrumumhüllende einer ChestResonance-Wellenform zeigt.

[0020] Fig. 8 ist ein Graph, der die Frequenzcharakteristiken der Resonanzen zeigt.

[0021] Fig. 9 ist ein Graph, der ein Beispiel von SpectralShapeDifference zeigt.

[0022] Fig. 10 ist ein Graph, der die Größenspektrumumhüllende der harmonischen Komponenten HC aus Fig. 2 zeigt, welche in EpR-Parameter zerlegt worden sind.

[0023] Fig. 11A und Fig. 11B sind Graphen, die Beispiele der Gesamtspektrumumhüllenden zeigen, wenn EGain von ExcitationCurve aus Fig. 10 geändert wird.

[0024] Fig. 12A und Fig. 12B sind Graphen, die Beispiele der Gesamtspektrumumhüllenden zeigen, wenn ESlope von ExcitationCurve aus Fig. 10 geändert wird.

[0025] Fig. 13A und Fig. 13B sind Graphen, die Beispiele der Gesamtspektrumumhüllenden zeigen, wenn ESlopeDepth von ExcitationCurve aus Fig. 10 geändert wird.

[0026] Fig. 14A bis Fig. 14C sind Graphen, welche eine Änderung in EpR zusammen mit einer Änderung in Dynamics zeigen.

[0027] Fig. 15 ist ein Graph, der eine Änderung in den Frequenzcharakteristiken zeigt, wenn Opening geändert wird.

[0028] Fig. 16 ist ein Blockdiagramm eines lied-synthetisierenden Antriebs bzw. Moduls einer Stimmensynthesevorrichtung.

Beschreibung der bevorzugten Ausführungsformen

[0029] Fig. 1 ist ein Diagramm, das eine Stimmenanalyse bzw. -zerlegung zeigt.

[0030] Stimmen, die in eine Stimmeneingabeeinheit **1** eingegeben werden, werden an eine Stimmenanalyse- bzw. zerlegeeinheit **2** gesendet. Die Stimmenanalyseeinheit **2** analysiert bzw. zerlegt die gelieferten Stimmen zu jedem konstanten Zeitraum. Die Stimmenanalyseeinheit **2** zerlegt eine Eingangsstimme in harmonische Komponenten HC und nicht-harmonische Komponenten US, beispielsweise durch Spektralmodulierungssynthese (SMS).

[0031] Die harmonischen Komponenten HC sind Komponenten, die mittels einer Summe von Sinuswellen mit bestimmten Frequenzen und Amplituden bzw. Größen dargestellt werden können. In Fig. 2 gezeigte Punkte deuten die Frequenz und Größe (Sinuskomponenten) einer Eingangsstimme an, die man als harmonische Komponenten HC erhält. In dieser Ausführungsform wird ein Satz gerader Linien, welche diese Punkte verbindet, als eine Größenspektrumumhüllende verwendet. Die Größenspektrumumhüllende ist in Fig. 2 als gestrichelte Linie gezeigt. Eine Basis- bzw. Grundfrequenz Pitch lässt sich zur gleichen Zeit erhalten, zu der man die harmonischen Komponenten HC erhält.

[0032] Die nicht-harmonischen Komponenten UC sind Rauschkomponenten der Eingangsstimme, die nicht in harmonische Komponenten HC zerlegt werden können. Die nicht-harmonischen Komponenten UC sind beispielsweise die in Fig. 3 gezeigten. Der obere Graph in Fig. 3 zeigt ein Größenspektrum, das repräsentativ für die Größe der nicht-harmonischen Komponenten UC ist, und der untere Graph zeigt ein Phasenspektrum, welches repräsentativ für die Phase der nicht-harmonischen Komponenten UC ist. In dieser Ausführungsform sind die Größen und Phasen der nicht-harmonischen Komponenten UC selbst als Rahmeninformation FL aufgezeichnet worden.

[0033] Die Größenspektrumumhüllende der durch die Analyse bzw. Zerlegung extrahierten harmonischen Komponenten wird in eine Vielzahl von Anregungs- plus Resonanz(EpR)-Parametern zerlegt, um spätere Verfahrensschritte zu erleichtern.

[0034] In dieser Ausführungsform umfassen die EpR-Parameter vier Parameter: einen ExcitationCurve (Anregungskurven)-Parameter, einen VocalTractResonance (Vokaltraktresonanz)-Parameter, einen ChestResonance (Brustresonanz)-Parameter und einen SpectralShapeDifferential (Spektralformunterschieds)-Parameter. Andere EpR-Parameter können ebenfalls verwendet werden.

[0035] Wie weiter unten ausgeführt, deutet die ExcitationCurve eine Spektrumumhüllende einer Stimmbandschwingungswellenform an, und die VocalTractResonance ist eine Näherung der Spektrumsform (Formanten), die in einem Vokaltrakt als eine Kombination verschiedener Resonanzen gebildet wird. Die ChestResonance ist, im Gegensatz zu den Formanten der VocalTractResonance, eine Näherung der Formanten niedriger Frequenz, welche als eine Kombination verschiedener Resonanzen (insbesondere Brustresonanzen) gebildet werden.

[0036] SpectralShapeDifferential stellt die Komponenten dar, welche sich nicht durch die oben beschriebenen drei EpR-Parameter ausdrücken lassen. Insbesondere erhält man SpectralShapeDifferential durch Subtrahieren von ExcitationCurve, VocalTractResonance und ChestResonance von der Größenspektrumumhüllenden.

[0037] Die nicht-harmonischen Komponenten UC und die EpR-Parameter werden in einer Speichereinheit **3** als Teile der Rahmeninformation FL1 bis FLn gespeichert.

[0038] Fig. 4 ist ein Graph, der die Spektrumumhüllende (ExcitationCurve) einer Stimmbandschwingungswellenform zeigt. Die ExcitationCurve korrespondiert zu der Größenspektrumumhüllenden einer Stimmbandschwingungswellenform.

[0039] Im speziellen ist die ExcitationCurve aus drei EpR-Parametern aufgebaut: aus einee EGain [dB] Darstellenden der Größe einer Stimmbandschwingungswellenform; einer ESlope-Darstellenden einer Steigung der Spektrumumhüllenden der Stimmbandschwingungswellenform; und aus einer ESlopeDepth-Darstellenden einer Tiefe bzw. eines Abstands vom Maximalwert zum Minimalwert der Spektrumumhüllenden der Stimmbandschwingungswellenform.

[0040] Durch Nutzung dieser drei EpR-Parameter kann die Größenspektrumumhüllende (ExcitationCurve Mag dB) der ExcitationCurve bei einer Frequenz fHz durch die folgende Gleichung dargestellt werden:

$$ExcitationCurveMag_{dB}(f_{HZ}) = EGain_{dB} + ESlopeDepth_{dB} \cdot (e^{-ESlope \cdot f_{HZ}} - 1) \qquad \text{(a)}$$

[0041] Aus dieser Gleichung (a) wird es verständlich, dass EGain die Signalgröße bzw. Signalstärke der Größenspektrumumhüllenden von ExcitationCurve tatsächlich ändern kann, und ESlope und ESlopeDepth können die Frequenzcharakteristiken (Steigung) der Signalhöhe der Größenspektrumumhüllenden von ExcitationCurve steuern.

[0042] Fig. 5 ist ein Graph, der eine Änderung in ExcitationCurve aus Gleichung (a) zeigt. ExcitationCurve bewegt sich, startend von EGain [dB] bei einer Frequenz f = 0 Hz, entlang einer Asymptote von EGain – ESlopeDepth [dB]. ESlope bestimmt die Steigung von ExcitationCurve.

[0043] Als Nächstes wird beschrieben, wie EGain, ESlope und ESlopeDepth berechnet werden. Durch Extrahieren bzw. Herauslösen der EpR-Parameter aus der Größenspektrumumhüllenden der ursprünglichen harmonischen Komponenten HC werden die ersten der oben beschriebenen drei EpR-Parameter berechnet.

[0044] Beispielsweise werden EGain, ESlope und ESlopeDepth nach der folgenden Methode berechnet.

[0045] Zuerst wird die maximale Größe der ursprünglichen harmonischen Komponenten HC bei einer Frequenz von 250 Hz oder weniger auf MAX [dB] gesetzt, und MIN wird auf –100[dB] gesetzt.

[0046] Als Nächstes werden die Größe und Frequenz der i-ten Sinuskomponente der ursprünglichen harmonischen Komponenten HC bei einer Frequenz von 10.000 Hz auf Sin Mag[1] [dB] und Sin Freq[i] [Hz] gesetzt, und die Zahl der Sinuskomponenten bei der Frequenz von 10000 Hz wird auf N gesetzt. Die Mittelwerte werden aus den folgenden Gleichungen (b1) und (b2) berechnet, wobei Sin Freq[0] die niedrigste Frequenz der Sinuskomponenten darstellt:

$$XAverage = \frac{\sum_{i=0}^{i=N-1}(SinFreq[i] - SinFreq[0])}{N} \qquad \text{... (b1)}$$

$$YAverage = \frac{\sum_{i=0}^{i=N-1}(\log(SinMag[i] - MIN))}{N} \qquad \text{... (b2)}$$

[0047] Durch Verwendung der Gleichungen (b1) und (b2) ergeben sich die folgenden Gleichungen:

$a = \log(MAX - MIN)$ (b3)

$b = (a - YAverage)/XAverage$ (b4)

$A = e^a$ (b5)

$$B = -b \tag{b6}$$

$$A0 = A \cdot e^{-B \cdot SinFreq[0]} \tag{b7}$$

[0048] Durch Nutzung der Gleichungen (b3) bis (b7) werden EGain, ESlope und ESlopeDepth mittels der folgenden Gleichungen (b8), (b9) und (b10) berechnet:

$$EGain = A0 + MIN \tag{b8}$$

$$ESlopeDepth = A0 \tag{b9}$$

$$ESlope = B \tag{b10}$$

[0049] Die EpR-Parameter aus EGain, ESlope und ESlopeDepth können in der oben beschriebenen Weise berechnet werden.

[0050] Fig. 6 ist ein Graph, der eine Spektrumumhüllende zeigt, welche durch VocalTractResonance gebildet wird. VocalTractResonance ist eine Näherung der Spektrumsform (Formanten), welche durch einen Vokaltrakt als eine Kombination verschiedener Resonanzen gebildet wird.

[0051] Beispielsweise korrespondiert ein Unterschied zwischen Phonemen so wie "a" und "i", die durch einen Menschen gebildet werden, mit einem Unterschied der Formen von Bergen einer Größenspektrumumhüllenden, welcher hauptsächlich durch eine Änderung in der Form bzw. im Aussehen des Vokaltrakts erzeugt wird. Dieser Berg wird als Formant bezeichnet. Eine Näherung der Formanten kann mittels einer Nutzung von Resonanzen erlangt werden.

[0052] In dem in Fig. 6 gezeigten Beispiel werden Formanten durch das Verwenden von 11 Resonanzen angenähert. Die i-te Resonanz wird durch Resonance[i] dargestellt, und die Größe der i-ten Resonanz bei einer Frequenz f wird durch Resonance[i] Mag(f) dargestellt bzw. repräsentiert. Die Größenspektrumumhüllende von VocalTractResonance lässt sich aus der folgenden Gleichung (c1) erhalten:

$$YocalTractResonannceMag_{dB}(f_{Hz}) = TodB(\sum_{i} Resonance[i]Mag_{linear}(f_{Hz})) \tag{c1}$$

[0053] Durch Darstellen der Phase der i-ten Resonanz durch Resonance[i]Phase[f], kann die Phase (Phasenspektrum) von VocalTractResonance durch die folgende Gleichung (c2) dargestellt werden:

$$VocalTractResonanncePhase(f_{Hz}) = \sum_{i} Resonance[i]Phase(f_{Hz}) \tag{c2}$$

[0054] Jede Resonance[i] kann durch drei EpR-Parameter ausgedrückt werden: eine Mittenfrequenz F, eine Bandbreite bzw. Bandweite Bw und eine Amplitude Amp. Wie eine Resonanz berechnet wird, wird weiter unten beschrieben.

[0055] Fig. 7 ist ein Graph, der eine Spektrumumhüllende (ChestResonance) einer Brustresonanzwellenform. ChestResonance wird mittels einer Brustresonanz gebildet und durch Berge (Formanten) der Größenspektrumumhüllenden bei kleinen Frequenzen dargestellt, welche nicht durch VocalTractResonance dargestellt werden können, wobei die Berge (Formanten) durch Nutzen von Resonanzen gebildet werden.

[0056] Die i-te Resonanz der Brustresonanzen wird durch CResonance[i] dargestellt bzw. repräsentiert, und die Größe der i-ten Resonanz bei einer Frequenz f wird durch CResonance[i] Mag(f) dargestellt. Die Größenspektrumumhüllende von ChestResonance kann durch die folgende Gleichung (d) gebildet werden:

$$ChestResonanceMag_{dB}(f_{Hz}) = TodB(\sum_{i} CResonance[Mag_{linear}(f_{Hz})) \tag{d}$$

[0057] Jede CResonance[i] kann durch drei EpR-Parameter ausgedrückt werden: eine Mittenfrequenz F, eine Bandbreite bzw. Bandweite Bw und eine Amplitude Amp. Wie eine Resonanz berechnet wird, wird weiter unten beschrieben.

[0058] Jede Resonanz (Resonance[i], CResonance[i] aus VocalTractResonance und ChestResonance) kann durch drei EpR-Parameter definiert werden: die Mittenfrequenz F, Bandbreite bzw. Bandweite Bw und Amplitude Amp.

[0059] Die Transferfunktion einer z-Fläche einer Resonanz, welche die Mittenfrequenz F und eine Bandbreite Bw aufweist, kann durch die folgende Gleichung (e1) ausgedrückt werden:

$$T(z) = \frac{A}{1 - Bz^{-1} - Cz^{-2}} \quad \text{... (e1)}$$

wobei:

$z = e^{j2\pi fT}$ (e2)

T = Samplingperiod (e3)

$C = -e^{-2\pi fT}$ (e4)

$B = 2e^{-2\pi fT}\cos(2\pi fT)$ (e5)

A = 1 – B – C (e6)

[0060] Diese Frequenzantwort kann durch die folgende Gleichung (e7) ausgedrückt werden:

$$T(f) = \frac{1 - B - C}{1 - B\cos(2\pi fT) - C\cos(4\pi fT) + j\left[B\sin(2\pi fT) + C\sin(4\pi fT)\right]} \quad \text{... (e7)}$$

[0061] Fig. 8 ist ein Graph, der Beispiele der Frequenzcharakteristiken von Resonanzen zeigt. In diesen Beispielen betrug die Resonanzmittenfrequenz F 1500 Hz, und die Bandbreite Bw und Amplitude Amp wurden geändert.

[0062] Wie in Fig. 8 gezeigt, wird die Amplitude |T(f)| bei einer Frequenz f, die der Mittenfrequenz F entspricht, maximal. Dieser Maximalwert ist die Resonanzamplitude Amp. Resonance (f) (linearer Wert) einer Resonanz mit Mittenfrequenz F, Bandbreite Bw und Amplitude Amp (linearer Wert), welche durch Gleichung (e7) dargestellt wird, kann durch die folgende Gleichung (e8) ausgedrückt werden:

$$Resonance(f_{Hz}) = \frac{Amp_{linear}}{|T(F_{Hz})|} \cdot T(f_{Hz}) \quad \text{... (e8)}$$

[0063] Die Größe der Resonanz bei einer Frequenz f kann dadurch mittels der folgenden Gleichung (e9), und die Phase kann mittels der folgenden Gleichung (e10) ausgedrückt werden.

$ResonanceMag_{linear}(f_{Hz}) = |Resonance(f_{Hz})|$ (e9)

$ResonancePhase(f_{Hz}) = \angle Resonance(f_{Hz})$ (e10)

[0064] Fig. 9 zeigt ein Beispiel von SpectralShapeDifferential. SpectralShapeDifferential korrespondiert zu den Komponenten der Größenspektrumumhüllenden der ursprünglichen Eingangsstimme, welche sich nicht durch ExcitationCurve, VocalTractResonance und ChestResonance ausdrücken lassen.

[0065] Durch Darstellen dieser Komponenten durch SpectralShapeDifferential Mag(f)[dB], wird der folgenden Gleichung (f) genügt:

$OrgMag_{dB}(f_{Hz}) = ExcitationCurveMag_{dB}(f_{Hz}) + ChestResonanceMag_{dB}(f_{Hz})$
$+ YocalTractResonanceMag_{dB}(f_{Hz}) + SpetralShapeDifferentialMag_{dB}(f_{Hz})$ (f)

[0066] Und zwar ist SpectralShapeDifferential eine Differenz zwischen den anderen EpR-Parametern und den ursprünglichen harmonischen Komponenten, wobei diese Differenz aus einem konstanten Frequenzinteivall berechnet wird. Beispielsweise wird die Differenz in einem 50 Hz-Intervall berechnet, und eine geradlinige Interpolation wird zwischen benachbarten Punkten durchgeführt.

[0067] Die Größenspektrumumhüllende der harmonischen Komponenten der ursprünglichen Eingangsstim-

me kann aus Gleichung (f) unter Nutzung der EpR-Parameter reproduziert werden.

[0068] Ungefähr die gleiche ursprüngliche Eingangsstimme kann dadurch wiedererlangt werden, dass nicht-harmonische Komponenten zur Größenspektrumumhüllenden der reproduzierten bzw. wiederhergestellten harmonischen Komponenten hinzuaddiert werden.

[0069] Fig. 10 ist ein Graph, der die Größenspektrumumhüllende der harmonischen Komponenten HC aus Fig. 2, zerlegt in EpR-Parameter zeigt.

[0070] Fig. 10 zeigt: die zu Resonanzen mit einer Mittenfrequenz höher als der zweite in Fig. 6 gezeigte Berg korrespondierende VocalTractResonance; ChestResonance, die zu der geringsten in Fig. 7 gezeigten Mittenfrequenz aufweisenden Resonanz korrespondiert; SpectralShapeDifterential, welche mittels einer gepunkteten Linie in Fig. 9 angedeutet ist; und ExcitationCurve, welche mittels einer fett-gestrichelten Linie angedeutet ist.

[0071] Die zu VocalTractResonance und ChestResonance korrespondierenden Resonanzen werden zur ExcitationCurve addiert. SpectralShapeDifferential weist einen Differenzwert von 0 zu ExcitationCurve auf.

[0072] Als Nächstes wird beschrieben, wie die Gesamtspektrumumhüllende sich ändert, wenn ExcitationCurve geändert wird.

[0073] Fig. 11A und Fig. 11B zeigen Beispiele der Gesamtspektrumumhüllenden, wenn EGain aus ExcitationCurve aus Fig. 10 geändert wird.

[0074] Wie in Fig. 11A gezeigt, wird die Gesamtzunahme (Größe) der Gesamtspektrumumhüllenden groß, wenn EGain groß gemacht wird. Da sich jedoch die Form der Spektrumumhüllenden nicht ändert, wird die Klangfarbe nicht geändert. Nur das Volumen kann daher klein gemacht werden.

[0075] Fig. 12A und Fig. 12B zeigen Beispiele der Gesamtspektrumumhüllenden, wenn ESlope aus ExcitationCurve aus Fig. 10 geändert wird.

[0076] Wie in Fig. 12A gezeigt, ändert sich, wenn ESlope vergrößert wird, die Form der Spektrumumhüllenden, so dass sich die Klangfarbe ändert, obwohl sich die Zunahme (Größe) der Gesamtspektrumumhüllenden nicht ändert. Durch Setzen von ESlope auf einen hohen Wert kann man die unklare Klangfarbe mit einem unterdrückten Hochfrequenzbereich erhalten.

[0077] Wie in **Fig.** 12 gezeigt, ändert sich, wenn ESlope klein gemacht wird, die Form der Spektrumumhüllenden, so dass sich die Klangfarbe ändert, obwohl sich der Zuwachs (Größe) der Gesamtspektrumumhüllenden nicht ändert. Durch Setzen von ESlope auf einen kleinen Wert, kann man helle Klangfarben mit einem verbesserten Hochfrequenzbereich erhalten.

[0078] Fig. 13A und Fig. 13B zeigen Beispiele der Gesamtspektrumumhüllenden, wenn ESlopeDepth aus ExcitationCurve aus Fig. 10 geändert wird.

[0079] Wenn, wie in Fig. 13A gezeigt, ESlopeDepth groß gemacht wird, ändert sich die Form bzw. das Aussehen der Spektrumumhüllenden, so dass sich die Klangfarbe ändert, obwohl der Zuwachs (Größe) der Gesamtspektrumumhüllenden sich nicht ändert. Durch Setzen von ESlopeDepth auf einen großen Wert, kann man die unklare Klangfarbe mit einem unterdrückten Hochfrequenzbereich erhalten.

[0080] Wenn, wie in Fig. 13B gezeigt, ESlopeDepth klein gemacht wird, ändert sich die Form der Spektrumumhüllenden, so dass sich die Klangfarbe ändert, obwohl sich der Zuwachs (Größe) der Gesamtspektrumumhüllenden nicht ändert. Durch Setzen von ESlopeDepth auf einen kleinen Wert, kann man helle Klangfarben mit einem verbesserten Hochfrequenzbereich erhalten.

[0081] Die Effekte, die sich aus dem Ändern von ESlope und ESlopeDepth ergeben, sind sehr ähnlich.

[0082] Als Nächstes wird ein Verfahren zum Simulieren einer Änderung in der Klangfarbe einer echten Stimme beschrieben, wenn EpR-Parameter geändert werden. Falls beispielsweise angenommen wird, dass ein-rahmige Phonem-Daten eines gesprochenen Lauts wie beispielsweise "a" durch die EpR-Parameter und Dynamics (das Volumen bzw. die Lautstärke der Stimmenproduktion) dargestellt werden, wird eine Änderung

in der Klangfarbe, welche durch Dynamics aus einer Echtstimmenerzeugung geändert werden soll, durch eine Änderung der EpR-Parameter simuliert. Allgemein unterdrückt eine Stimmerzeugung bei einer geringen Lautstärke die Hochfrequenzkomponenten, und je größer die Lautstärke wird, desto stärker erhöhen sich die Hochfrequenzkomponenten, obwohl sich dies von einem zum anderen Stimmerzeuger ändert.

[0083] Fig. 14A bis Fig. 14C sind Graphen, die eine Änderung in den EpR-Parametern zeigen, wenn Dynamics geändert wird. Fig. 14A zeigt eine Änderung in EGain, Fig. 14B zeigt eine Änderung in ESlope und Fig. 14C zeigt eine Änderung in ESlopeDepth.

[0084] Die Abszissen in den Fig. 14A bis Fig. 14C repräsentieren einen Wert von Dynamics von 0 bis 1,0. Der Wert 0 von Dynamics repräsentiert die kleinste Stimmerzeugung, der Dynamics-Wert 1,0 repräsentiert die größte Stimmproduktion und der Dynamics-Wert 0,5 repräsentiert eine normale Stimmproduktion.

[0085] Eine Datenbank Timbre DB, welche weiter unten beschrieben wird, speichert EGain, ESlope und ESlopeDepth für eine normale Stimmerzeugung, diese EpR-Parameter werden in Übereinstimmung mit den in den Fig. 14A bis Fig. 14C gezeigten Funktionen geändert. Im speziellen wird die in Fig. 14A gezeigte Funktion durch FEGain (Dynamics) repräsentiert, die in Fig. 14B gezeigte Funktion wird durch FESlope (Dynamics) repräsentiert und die in Fig. 14C gezeigte Funktion wird durch FESlopeDepth (Dynamics) repräsentiert. Falls ein Dynamics-Parameter gegeben ist, können die Parameter durch die folgenden Gleichungen (g1) bis (g3) ausgedrückt werden:

$$NewEGain_{dB} = FEGain_{dB}(Dynamics) \quad (g1)$$

$$NewEslope = OriginalESlope*FESlope(Dynamics) \quad (g2)$$

$$NewESlopeDepth_{dB} = OriginalESlopeDept_{dB} + FESlopeDepth_{dB}(Dynamics) \quad (g3)$$

wobei Original ESlope und Original ESlopeDepth die ursprünglichen in der Datenbank Timbre DB gespeicherten EpR-Parameter sind.

[0086] Die in den Fig. 14A bis Fig. 14C gezeigten Funktionen erhält man durch Analysieren bzw. Zerlegen der Parameter der gleichen Phoneme, welche bei verschiedenen Graden bzw. Stärken der Stimmerzeugung (Dynamics) erzeugt werden. Durch Nutzen dieser Funktionen werden die EpR-Parameter in Übereinstimmung mit Dynamics geändert. Man kann berücksichtigen, dass die in den Fig. 14A bis Fig. 14C gezeigten Änderungen für jedes Phonem, jeden Stimmerzeuger usw. unterschiedlich sein können. Daher kann man durch Herstellen bzw. Anpassen der Funktion für jedes Phonem und jeden Stimmerzeuger eine Änderung erhalten, die analog zu einer realistischeren Stimmerzeugung ist.

[0087] Als Nächstes wird mit Bezug auf Fig. 15 ein Verfahren zum Reproduzieren einer Änderung in einer Klangfarbe beschrieben, wenn Opening eines Mundes bzw. eine Mundöffnung zur Stimmerzeugung des gleichen Phonems verändert wird.

[0088] Fig. 15 ist ein Graph, der eine Änderung in Frequenzcharakteristiken zeigt, wenn Opening geändert wird. Ähnlich zu Dynamics wird angenommen, dass der Opening-Parameter Werte von 0 bis 1,0 annimmt.

[0089] Der Opening-Wert 0 repräsentiert das kleinste Öffnen eines Mundes (niedriges Öffnen), der Opening-Wert 1,0 repräsentiert das größte Öffnen eines Mundes (hohes Öffnen) und der Opening-Wert 0,5 repräsentiert ein normales Öffnen eines Mundes (normales Öffnen).

[0090] Die später beschriebene Datenbank Timbre DB speichert EpR-Parameter ab, welche man erhält, wenn eine Stimme bei einer normalen Mundöffnung erzeugt wird. Die EpR-Parameter werden verändert, so dass sie die in Fig. 15 gezeigten Frequenzcharakteristiken bei dem gewünschten Grad an Mundöffnung zeigen.

[0091] Um diese Änderung zu realisieren, wird die Amplitude (EpR-Parameter) jeder Resonanz wie in Fig. 15 gezeigt. Beispielsweise werden die Frequenzcharakteristiken nicht geändert, wenn eine Stimme bei einem normalen Grad einer Mundöffnung (normales Öffnen) erzeugt wird. Wenn eine Stimme bei dem kleinsten Grad an Mundöffnung (niedriges Öffnen) erzeugt wird, werden die Amplituden der Komponenten bei 1 bis 5 kHz abgesenkt. Wenn eine Stimme beim größten Grad an Mundöffnung (hohe Öffnung) erzeugt wird, werden die Amplituden der Komponenten bei 1 bis 5 kHz angehoben.

[0092] Diese Änderungsfunktion wird durch FOpening (f) repräsentiert. Die EpR-Parameter können geändert werden, so dass sie die Frequenzcharakteristiken beim gewünschten Grad der Mundöffnung, d. h. bei den in Fig. 15 gezeigten Frequenzcharakteristiken aufweisen, und zwar durch Ändern der Amplitude jeder Resonanz nach der folgenden Gleichung (h):

$$\text{NewResonance[i]Amp}_{dB} = \text{OriginalResonance[i]Amp}_{dB} + \text{FOpening}_{dB}(\text{OriginalResonance[i]Freq}_{Hz})\cdot(0.5 - \text{Opening})/0.5 \quad \text{(h)}$$

[0093] Die Funktion FOpening von (f) erhält man durch Analysieren bzw. Zerlegen der Parameter der bei unterschiedlichen Graden der Mundöffnung erzeugten gleichen Phoneme. Durch Nutzen dieser Funktion werden die EpR-Parameter in Übereinstimmung mit den Opening-Werten geändert. Man kann berücksichtigen, dass sich diese Änderung für jedes Phonem, jeden Stimmerzeuger usw. ändern kann. Daher kann man durch Erstellen der Funktion für jedes Phonem und jeden Stimmerzeuger eine Änderung erreichen, die analog zu einer realistischeren Stimmerzeugung ist.

[0094] Die Gleichung (h) korrespondiert mit der i-ten Resonanz. Original Resonance[i]Amp und Original Resonance[i]Freq repräsentieren jeweils die Amplitude und Mittenfrequenz (EpR-Parameter) der in der Datenbank Timbre DB gespeicherten Resonanz. New Resonance[i]Amp repräsentiert die Amplitude einer neuen Resonanz.

[0095] Als Nächstes wird mit Bezug auf Fig. 16 beschrieben, wie ein Lied synthetisiert wird.

[0096] Fig. 16 ist ein Blockdiagramm eines lied-synthetisierenden Kerns bzw. Moduls einer Stimmensynthesevorrichtung. Das Lied-synthetisierende bzw. das das Lied künstlich herstellende Modul hat mindestens eine Eingabeeinheit **4**, eine Pulserzeugereinheit **5**, eine Fensterungs- & FFT ("FFT" = Fast Fourier Transformation)-Einheit **6**, eine Datenbank **7**, eine Vielzahl von Hinzufügungs- bzw. Additionseinheiten **8a** bis **8g** und eine IFFT (Inverse Fast Fourier Transformation)- & Überlappeinheit **9**.

[0097] In die Eingabeeinheit **4** werden eine Tonhöhe, eine Stimmintensität, eine Phonem- und andere Informationen in Übereinstimmung mit einer Melodie eines von einem Sänger gesungenen Liedes eingegeben, und zwar zu jeder Rahmen- bzw. Frame-Dauer, beispielsweise 5 ms. Die weitere Information ist beispielsweise eine Vibrato-Information einschließlich Vibratogeschwindigkeit und -tiefe. Die Informationseingabe in die Eingabeeinheit **4** wird in zwei Serien aufgeteilt, die zu der Pulserzeugereinheit **5** und der Datenbank **7** gesendet werden.

[0098] Die Pulserzeugereinheit **5** erzeugt auf der Zeitachse Pulse, welche ein Tonhöhen-Intervall aufweisen, welches zu einer Tonhöhen-Eingabe von der Eingabeeinheit **4** korrespondiert. Durch Ändern der Steigung und des Tonhöhen-Intervalls (Pitch-Intervalls) der erzeugten Pulse, um die erzeugten Pulse selbst mit einer Fluktuation bzw. Schwankung in der Steigung und dem Tonhöhen-Intervall zu versehen, können sogenannte rauhe bzw. barsche Stimmen und ähnliches erzeugt werden.

[0099] Falls der gerade vorliegende Frame bzw. Datenblock bzw. Rahmen ein stimmloser Laut ist, gibt es keine Tonhöhe, so dass das von der Pulserzeugereinheit **5** angewandte Verfahren nicht notwendig ist. Das von der Pulserzeugereinheit **5** angewandte Verfahren wird nur durchgeführt, wenn ein stimmlicher Laut erzeugt wird.

[0100] Die Fensterungs- & FFT-Einheit **6** erzeugt ein Fenster in einem Puls (Zeit-Wellenform), der durch die Pulserzeugereinheit **5** erzeugt wird, und führt dann eine schnelle Fourier-Transformation durch, um den Puls in eine Frequenzband-Information umzusetzen. Ein Größenspektrum der umgewandelten Frequenzband-Information ist über den gesamten Bereich flach. Eine Ausgabe von der Fensterungs- & FFT-Einheit **6** wird in das Phasenspektrum und Größen- bzw. Amplituden-Spektrum aufgeteilt.

[0101] Die Datenbank **7** bereitet mehrere Datenbanken vor, damit diese zum Synthetisieren von Stimmen eines Liedes verwendet werden. In dieser Ausführungsform bereitet die Datenbank **7** vor: Timbre DB, Stationary DB, Articulation DB, Note DB und Vibrato DB.

[0102] In Übereinstimmung mit der Informationseingabe in die Eingabeeinheit **4** liest die Datenbank **7** die notwendigen Datenbanken zur Berechnung der EpR-Parameter und die notwendigen nicht- bzw. anharmonischen

Komponenten für die Synthese zu einigen Zeitpunkten aus. Timbre DB speichert typische EpR-Parameter eines Frames bzw. Datenblocks für jedes Phonem eines stimmlichen Lauts (Vokal, Nasallaut, stimmlicher Konsonant). Sie speichert auch EpR-Parameter eines Frames des gleichen Phonems, die jede zu einer Mehrzahl von Tonhöhen korrespondieren. Durch Nutzen dieser Tonhöhen und Interpolation kann man die EpR-Parameter, die zu einer gewünschten Tonhöhe korrespondieren, erhalten.

[0103] Stationary DB speichert stabile Analyse-Frames aus mehreren Sekunden für jeden der erzeugten Phoneme in einer anhaltenden Art, als auch die harmonischen Komponenten (EpR-Parameter) und nicht-harmonischen Komponenten. Wird beispielsweise angenommen, dass die Frame-Dauer bzw. das Frame-Intervall 5 ms beträgt und die stabile Stimmerzeugungszeit 1 s beträgt, speichert Stationary DB die Information aus 200 Frames für jedes Phonem.

[0104] Da Stationary DB EpR-Parameter speichert, welche man durch Analyse bzw. Zerlegen einer Originalstimme erhalten hat, hat sie Informationen wie beispielsweise feine Fluktuationen bzw. Schwankungen der Originalstimme. Durch Nutzen dieser Information kann man feine Änderungen auf die EpR-Parameter aufgeben, welche man aus Timbre DB erhalten hat. Es ist daher möglich, die natürliche Tonlage, Anstieg, Resonanz usw. der Originalstimme zu reproduzieren. Durch Hinzufügen nicht-harmonischer Komponenten können noch natürlichere synthetisierte Stimmen realisiert werden.

[0105] Articulation speichert einen analysierten Änderungsteil von einem Phonem zu einem anderen Phonem als auch die harmonischen Komponenten (EpR-Parameter) und nicht-harmonischen Komponenten. Wenn eine Stimme synthetisiert wird, welche von einem Phonem zu einem anderen Phonem wechselt, wird auf Articulation verwiesen, und eine Änderung in den EpR-Parametern und den nicht-harmonischen Komponenten wird für diesen sich ändernden Teil dazu benutzt, um einen natürlichen Phonem-Wechsel zu reproduzieren.

[0106] Note DB ist aus drei Datenbanken: Attack DB, Release DB und Note Transition DB aufgebaut. Sie speichern Information einer Änderung im Zuwachs (EGain) und Tonhöhe und andere Information, welche durch eine Analyse bzw. Zerlegung der Originalstimme (Echtstimme) erhalten wurden, jeweils für einen Laut-Erzeugungs-Anfangsteil, einen Stimm-Auslass-Teil und einen Tonübergangsteil.

[0107] Falls beispielsweise eine Änderung im Zuwachs (EGain) und Tonlage, die in Attack DB gespeichert sind, zu den EpR-Parametern für den Lauterzeugungs-Anfangsteil hinzuaddiert werden, kann die Änderung im Zuwachs und Tonhöhe wie eine natürliche Echtstimme zu der synthetisierten Stimme hinzugefügt werden.

[0108] Vibrato DB speichert Information über eine Änderung in Zuwachs (EGain) und Tonhöhe und andere Information, welche durch eine Analyse eines Vibrato-Teils der Originalstimme (Echtstimme) erhalten wurde.

[0109] Falls beispielsweise ein Vibrato-Teil existiert, der zu einer zu synthetisierenden Stimme hinzugegeben werden soll, werden EpR-Parameter des Vibrato-Teils hinzugefügt, und zwar mit einer in Vibrato DB gespeicherten Änderung in Zuwachs (EGain) und Tonhöhe, so dass eine natürliche Änderung in Zuwachs und Tonhöhe zur synthetisierten Stimme hinzugefügt werden kann. Und zwar kann ein natürliches Vibrato reproduziert werden.

[0110] Obwohl diese Ausführungsformen fünf Datenbanken vorsieht, kann die künstliche Erzeugung bzw. Synthese von Stimmen eines Lieds grundsätzlich unter Verwendung mindestens von Timbre DB, Stationary DB und Articulation DB durchgeführt werden, falls die Information über Stimmen eines Liedes und Tonhöhen, Stimmenlautstärken und Mundöffnungsgrade gegeben ist.

[0111] Stimmen eines in Ausdruck reichen Liedes können unter Nutzung der zusätzlichen Datenbanken Note DB und Vibrato DB synthetisiert werden. Die hinzufügbaren Datenbanken sind nicht nur auf Note DB und Vibrato DB beschränkt, sondern es kann jede Datenbank für einen Stimmausdruck verwendet werden.

[0112] Die Datenbank **7** gibt die EpR-Parameter von ExcitationCurve EC, ChestResonance CR, VocalTractResonance VTR und SpectralShapeDifferential SSD aus, welche durch Nutzung der oben beschriebenen Datenbank berechnet worden sind, und weiterhin die nicht-harmonischen Komponenten UC.

[0113] Als nicht-harmonische Komponenten UC gibt die Datenbank **7** das Größenspektrum und Phasenspektrum, so wie in **Fig. 3** gezeigt, aus. Die nicht-harmonischen Komponenten US repräsentieren Rauschkomponenten eines stimmhaften Lauts der Originalstimme, welcher sich nicht als harmonische Komponenten ausdrücken lässt, und einesn stimmlosen Lauts, der sich inhärent nicht als harmonische Komponente ausdrücken

lässt.

[0114] Wie in Fig. 16 gezeigt, werden VocalTractResonance VTR und nicht-harmonische Komponenten getrennt für Phase und Größe bzw. Amplitude ausgegeben.

[0115] Die Additionseinheit **8a** fügt ExcitationCurve EC zur Ausgabe des flachen Größenspektrums der Fensterungs- & FFT-Einheit **6** hinzu. Und zwar wird die Größe bei jeder Frequenz, die durch die Gleichung (a) unter Nutzung von EGain, ESlope und ESlopeDepth berechnet worden ist, hinzuaddiert. Das Additionsergebnis wird in einem folgenden Schritt zur Additionseinheit **8b** gesandt.

[0116] Das erhaltene Größenspektrum ist eine Größenspektrumumhüllende (Excitation Curve) einer Vokaltrakt-Schwingungs-Wellenform, so wie sie in Fig. 4 gezeigt ist.

[0117] Durch Ändern von EGain, ESloe und ESlopeDepth in Übereinstimmung mit den in den Fig. 14A bis Fig. 14C gezeigten Funktionen unter Verwendung der Dynamics-Parameter lässt sich eine Änderung in der Klangfarbe ausdrücken, die durch eine Änderung in der Stimmlautstärke erzeugt wird.

[0118] Falls die Stimmlautstärke geändert werden soll, wird EGain wie in den Fig. 11A und Fig. 11B gezeigt, geändert. Falls die Klangfarbe geändert werden soll, wird ESlope, wie in den Fig. 12A und Fig. 12B, geändert.

[0119] Die Additionseinheit **8b** fügt ChestResonance CR, welches durch Gleichung (d) erhalten wurde, zum Größenspektrum hinzu, dem ExcitationCurve EC in der Additionseinheit **8a** hinzugefügt worden ist, um so das Größenspektrum zu erhalten, dem der Berg des Größenspektrums der Brustresonanz, so wie in Fig. 7 gezeigt, hinzugefügt worden ist. Das erhaltene Größenspektrum wird in einem weiteren Schritt zur Additionseinheit **8c** gesendet.

[0120] Indem man die Größe von ChestResonance CR groß macht, ist es möglich, den Brustresonanzlaut größer als bei der ursprünglichen Stimmqualität einzustellen. Durch Erniedrigung der Frequenz von ChestResonance CR ist es möglich, die Stimme so zu ändern, dass die Stimme einen niedrigeren Brustresonanz-Laut aufweist.

[0121] Die Additionseinheit **8c** fügt VocalTractResonance VTR, das aus Gleichung (c1) erhalten wurde, zum Größenspektrum hinzu, dem ChestResonance CR in der Additionseinheit **8b** hinzugefügt worden ist, um so das Größenspektrum zu erhalten, dem der Berg des größten Spektrums des Vokaltrakts, so wie in Fig. 6 gezeigt, hinzugefügt wurde. Das erhaltene Größenspektrum wird in einem weiteren Schritt zur Additionseinheit **8e** gesandt.

[0122] Durch Hinzufügen von VocalTractResonance VTR ist grundsätzlich möglich, einen Unterschied zwischen Klangfarben, die durch einen Unterschied zwischen Phonemen, wie beispielsweise "a" und "i", erzeugt werden, auszudrücken.

[0123] Durch Ändern der Amplitude jeder Resonanz in Überstimmung mit dem in Fig. 15 beschriebenen Opening-Parameter unter Nutzung der Frequenzfunktion kann eine durch einen Grad einer Mundöffnung erzeugte Änderung in der Klangfarbe reproduziert werden.

[0124] Durch Ändern der Frequenz, Größe und Bandbreite jeder Resonanz kann die Lautqualität hin zu einer Lautqualität geändert werden, welche unterschiedlich von der ursprünglichen Lautqualität ist (beispielsweise zur Lautqualität einer Oper). Durch Ändern der Tonhöhe können männliche Stimmen in weibliche Stimmen umgewandelt werden, oder umgekehrt. Die Additionseinheit **8d** fügt VocalTractResonance VTR, das durch Gleichung (c2) erhalten wurde, zur Ausgabe des flachen Phasenspektrums aus der Fensterungs- & FFT-Einheit **6** hinzu. Das erhaltene Phasenspektrum wird zur Additionseinheit **8g** gesandt.

[0125] Die Additionseinheit **8e** fügt SpectralShapeDifferential Mag dB (fHz) zu dem Größenspektrum, zu dem VocalTractResonance VTR an der Additionseinheit **8c** hinzugefügt wird, hinzu, um ein präziseres Größenspektrum zu erhalten.

[0126] Die Additionseinheit **8f** addiert das Größenspektrum der nicht-harmonischen Komponenten UC, das von der Datenbank **7** geliefert wird, und das Größenspektrum, das von der Additionseinheit **8e** gesandt wurde, zusammen. Das zusammenaddierte Größenspektrum wird in einem folgenden Schritt zur IFFT- & Überlapp-Additionseinheit **9** weitergeleitet.

[0127] Die Additionseinheit **8g** addiert das von der Datenbank **7** gelieferte Phasenspektrum der nicht-harmonischen Komponenten und das von der Additionseinheit **8d** gelieferte Phasenspektrum zusammen. Das aufaddierte Phasenspektrum wird zur IFFT- & Überlapp-Additionseinheit **9** gesandt.

[0128] Die IFFT- & Überlapp-Additionseinheit **9** führt eine inverse Fast Fourier-Transformation des lieferten Größenspektrums und Phasenspektrums durch, und fügt überlappend die transformierten Zeit-Wellenformen zusammen, um die endgültigen synthetisierten Stimmen zu erzeugen.

[0129] Gemäß der Ausführungsform wird eine Stimme in harmonische Komponenten und nicht-harmonische Komponenten zerlegt. Die analysierten bzw. zerlegten harmonischen Komponenten können in die Größenspektrumumhüllende und eine Vielzahl von Resonanzen jeweils einer Stimmband-Wellenform zerlegt werden, und in einen Unterschied zwischen diesen Umhüllenden und Resonanzen und der Originalstimme, welche gespeichert werden.

[0130] Gemäß der Ausführungsform kann die Größenspektrumumhüllende einer Stimmbandwellenform durch drei EpR-Parameter EGain, ESlope und ESlope-Depth repräsentiert werden.

[0131] Gemäß der Ausführungsform kann durch Ändern der EpR-Parameter, die zu einer Änderung in der Stimm-Lautstärke in Übereinstimmung mit einer vorbestimmten Funktion korrespondieren, eine Stimme mit einer natürlichen Änderung der Klangfarbe, welche durch eine Änderung in der Lautstärke erzeugt wird, synthetisiert werden.

[0132] Gemäß der Ausführungsform kann durch Ändern der EpR-Parameter, die zu einer Änderung im Grad der Mundöffnung in Übereinstimmung mit einer vorbestimmten Funktion korrespondieren, eine Stimme synthetisiert werden, bei der eine natürliche Änderung der Klangfarbe durch eine Änderung im Grad der Mundöffnung erzeugt wird.

[0133] Da sich die Funktionen mit jedem Phonem und jedem Stimmenerzeuger ändern können, kann eine Stimme synthetisiert werden, indem eine individuelle charakteristische Differenz zwischen Änderungen in der Klangfarbe, die durch Phoneme und Stimmerzeuger erzeugt wird, berücksichtigt wird.

[0134] Obwohl die Ausführungsform hauptsächlich mit Bezug auf die Erzeugung von Stimmen eines durch eines Sängers gesungenen Liedes beschrieben wird, ist die Ausführungsform nicht darauf beschränkt, sondern allgemeine Sprachlaute und Musikinstrument-Laute können ebenfalls in einer gleichen Art synthetisiert werden.

[0135] Die Ausführungsform kann durch einen Computer und dergleichen realisiert werden, welcher mit einem Computerprogramm usw. ausgerüstet ist, welches die dargestellten Funktionen realisiert. In diesem Falle kann das Computerprogramm und dergleichen, das die dargestellten Funktionen realisiert, in einem Computer-lesbaren Speichermedium, wie beispielsweise einer CD-ROM und einer Floppy Disc gespeichert werden, um zu einem Anwender verschickt zu werden.

[0136] einer Diskette gespeichert werden, um zu einem Anwender verschickt zu werden.

[0137] Falls der Computer und dergleichen mit einem Kommunikations-Netzwerk, wie beispielsweise einem LAN, dem Internet und einer Telefonleitung verbunden ist, können das Computerprogramm, Daten usw. über das Kommunikations-Netzwerk verbreitet werden.

[0138] Die vorliegende Erfindung, so wie sie in den Ansprüchen beansprucht wird, ist in Verbindung mit den bevorzugten Ausführungsformen beschrieben worden. Die Erfindung ist nicht nur auf die oben beschriebenen Ausführungsformen beschränkt. (TK-E-19450) Es ist offensichtlich, dass verschiedene Modifikationen, Verbesserungen, Kombinationen usw. durch den Fachmann durchgeführt werden können.

Patentansprüche

1. Stimmenanalysevorrichtung, die Folgendes aufweist:
erste Analysemittel (**2**) zum Analysieren bzw. Zerlegen einer Stimme in harmonische Komponenten und nicht-harmonische Komponenten;
zweite Analysemittel zum Analysieren bzw. Zerlegen einer Größenspektrumumhüllenden der harmonischen Komponenten in eine Größenspektrumumhüllende einer Stimmbandschwingungswellenform, Resonanzen so-

wie eine Spektrumumhüllende einer Differenz zwischen der Größenspektrumumhüllenden der harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen; und
Mittel (**3**) zum Speichern der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen sowie der Spektrumumhüllenden der Differenz.

2. Stimmenanalysevorrichtung gemäß Anspruch 1, wobei:
die Größenspektrumumhüllende der Stimmbandschwingungswellenform repräsentiert ist durch drei Parameter EGain, ESlope und ESlopeDepth; und
die drei Parameter ausgedrückt werden können durch die folgende Gleichung (1):

$$\text{ExcitationCurveMag}(f) = \text{EGain} + \text{ESlopeDepth}\cdot(e^{-\text{ESlope}\cdot f} - 1) \quad (1)$$

wobei ExcitationCurveMag(f) die Größenspektrumumhüllende der Stimmbandschwingungswellenform ist.

3. Stimmenanalysevorrichtung gemäß Anspruch 1, wobei die Resonanzen eine Vielzahl von Resonanzen umfassen, die Vokaltraktformanten ausdrücken, sowie eine Resonanz umfassen, die Brustresonanz ausdrückt.

4. Stimmensynthesevorrichtung, die Folgendes aufweist:
Mittel (**7**) zum Speichern von nicht-harmonischen Komponenten, die aus einer Stimme analysiert wurden, von einer Größenspektrumumhüllenden einer Stimmbandschwingungswellenform, von Resonanzen sowie von einer Spektrumumhüllenden einer Differenz zwischen einer Größenspektrumumhüllenden von harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen, wobei die Größenspektrumumhüllende, die Resonanzen und die Spektrumumhüllende der Differenz aus den harmonischen Komponenten analysiert wurden, welche aus der Stimme analysiert wurden;
Mittel (**4**) zur Eingabe von Information über eine zu synthetisierende Stimme;
Mittel (**6**) zum Erzeugen einer flachen Größenspektrumumhüllenden; und
Mittel (**8**) zum Hinzufügen der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden für die Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz, die jeweils aus den Mitteln zum Speichern ausgelesen wurden, zu der flachen Größenspektrumumhüllenden in Übereinstimmung mit der eingegebenen Information.

5. Stimmensynthesevorrichtung gemäß Anspruch 4, wobei:
die Größenspektrumumhüllende der Stimmbandschwingungswellenform repräsentiert ist durch drei Parameter EGain, ESlope und ESlopeDepth; und
die drei Parameter ausgedrückt werden können durch die folgende Gleichung (1):

$$\text{ExcitationCurveMag}(f) = \text{EGain} + \text{ESlopeDepth}\cdot(e^{-\text{ESlope}\cdot f} - 1) \quad (1)$$

wobei ExcitationCurveMag(f) die Größenspektrumumhüllende der Stimmbandschwingungswellenform ist.

6. Stimmensynthesevorrichtung gemäß Anspruch 5, wobei die Mittel zum Speichern ferner eine Funktion zum Ändern der drei Parameter speichern, und zwar in Übereinstimmung mit einer Änderung des Klangvolumens bzw. der Lautstärke, so dass die Klangfarbe verändert werden kann in Übereinstimmung mit der Änderung des Klangvolumens bzw. der Lautstärke.

7. Stimmensynthesevorrichtung gemäß Anspruch 4, wobei die Resonanzen eine Vielzahl von Resonanzen umfassen, die Vokaltraktformanten ausdrücken, sowie eine Resonanz umfassen, die Brustresonanz ausdrückt.

8. Stimmensynthesevorrichtung gemäß Anspruch 7, wobei die Mittel zum Speichern ferner eine Funktion zum Ändern einer Amplitude jeder Resonanz speichern, und zwar in Übereinstimmung mit einem Mundöffnungsgrad, so dass die Klangfarbe verändert werden kann in Übereinstimmung mit dem Mundöffnungsgrad.

9. Stimmenanalyse- und -synthesevorrichtung, die Folgendes aufweist:
erste Analysemittel zum Analysieren bzw. Zerlegen einer Stimme in harmonische Komponenten und nicht-harmonische Komponenten;
zweite Analysemittel zum Analysieren bzw. Zerlegen einer Größenspektrumumhüllenden der harmonischen

Komponenten in eine Größenspektrumumhüllende einer Stimmbandschwingungswellenform, Resonanzen sowie eine Spektrumumhüllende einer Differenz zwischen der Größenspektrumumhüllenden der harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen;
Mittel zum Speichern der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen sowie der Spektrumumhüllenden der Differenz;
Mittel zur Eingabe von Information über eine zu synthetisierende Stimme;
Mittel zum Erzeugen einer flachen Größenspektrumumhüllenden; und
Mittel zum Hinzufügen der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden für die Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz, die jeweils aus den Mitteln zum Speichern ausgelesen wurden, zu der flachen Größenspektrumumhüllenden in Übereinstimmung mit der eingegebenen Information.

10. Stimmenanalyseverfahren, das die folgenden Schritte aufweist:
(a) Analysieren bzw. Zerlegen einer Stimme in harmonische und nicht-harmonische Komponenten;
(b) Analysieren bzw. Zerlegen einer Größenspektrumumhüllenden der harmonischen Komponenten in eine Größenspektrumumhüllende einer Stimmbandschwingungswellenform, Resonanzen und eine Spektrumumhüllende einer Differenz zwischen der Größenspektrumumhüllenden der harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen; und
(c) Speichern der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz.

11. Stimmensyntheseverfahren, das die folgenden Schritte aufweist:
(a) Auslesen nicht-harmonischer Komponenten, die aus einer Stimme analysiert wurden, einer Größenspektrumumhüllenden einer Stimmbandschwingungswellenform, von Resonanzen und einer Spektrumumhüllenden einer Differenz zwischen einer Größenspektrumumhüllenden von harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen, wobei die Größenspektrumumhüllende, die Resonanzen und die Spektrumumhüllende einer Differenz aus den harmonischen Komponenten analysiert wurden, welche aus der Stimme analysiert wurden;
(b) Eingabe von Information über eine zu synthetisierende Stimme;
(c) Erzeugen einer flachen Größenspektrumumhüllenden; und
(d) Hinzufügen der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz, die jeweils im Schritt (a) ausgelesen wurden, zu der flachen Größenspektrumumhüllenden in Übereinstimmung mit der eingegebenen Information.

12. Ein Programm, welches ein Computer ausführt zum Realisieren eines Musikdatenspielprozesses, wobei das Programm die folgenden Instruktionen aufweist:
(a) Analysieren bzw. Zerlegen einer Stimme in harmonische und nicht-harmonische Komponenten;
(b) Analysieren bzw. Zerlegen einer Größenspektrumumhüllenden der harmonischen Komponenten in eine Größenspektrumumhüllende einer Stimmbandschwingungswellenform, Resonanzen und eine Spektrumumhüllende einer Differenz zwischen der Größenspektrumumhüllenden der harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen; und
(c) Speichern der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz.

13. Ein Programm, welches ein Computer ausführt zum Realisieren eines Musikdatenspielprozesses, wobei das Programm die folgenden Instruktionen aufweist:
(a) Auslesen nicht-harmonischer Komponenten, die aus einer Stimme analysiert wurden, einer Größenspektrumumhüllenden einer Stimmbandschwingungswellenform, von Resonanzen und einer Spektrumumhüllenden einer Differenz zwischen einer Größenspektrumumhüllenden von harmonischen Komponenten und einer Summe der Größenspektrumumhüllenden der Stimmbandschwingungswellenform und der Resonanzen, wobei die Größenspektrumumhüllende, die Resonanzen und die Spektrumumhüllende einer Differenz aus den harmonischen Komponenten analysiert wurden, welche aus der Stimme analysiert wurden;
(b) Eingabe von Information über eine zu synthetisierende Stimme;
(c) Erzeugen einer flachen Größenspektrumumhüllenden; und
(d) Hinzufügen der nicht-harmonischen Komponenten, der Größenspektrumumhüllenden der Stimmbandschwingungswellenform, der Resonanzen und der Spektrumumhüllenden der Differenz, die jeweils im Schritt

(a) ausgelesen wurden, zu der flachen Größenspektrumumhüllenden in Übereinstimmung mit der eingegebenen Information.

Es folgen 11 Blatt Zeichnungen

Anhängende Zeichnungen

FIG. 1

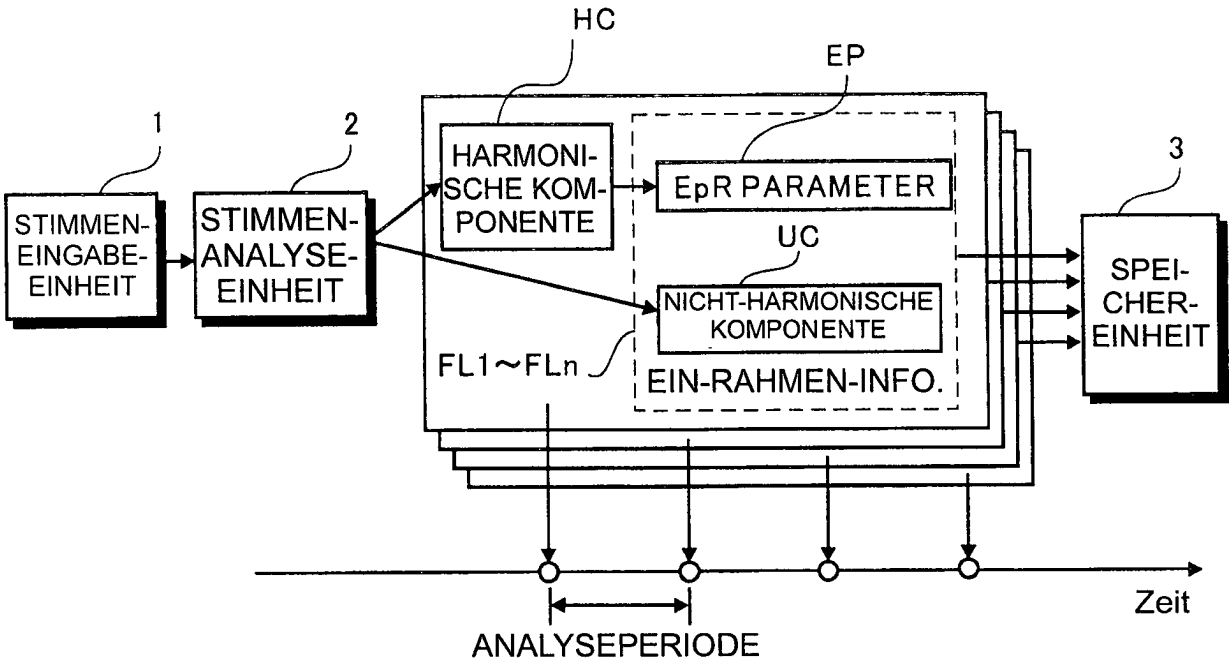


FIG. 2

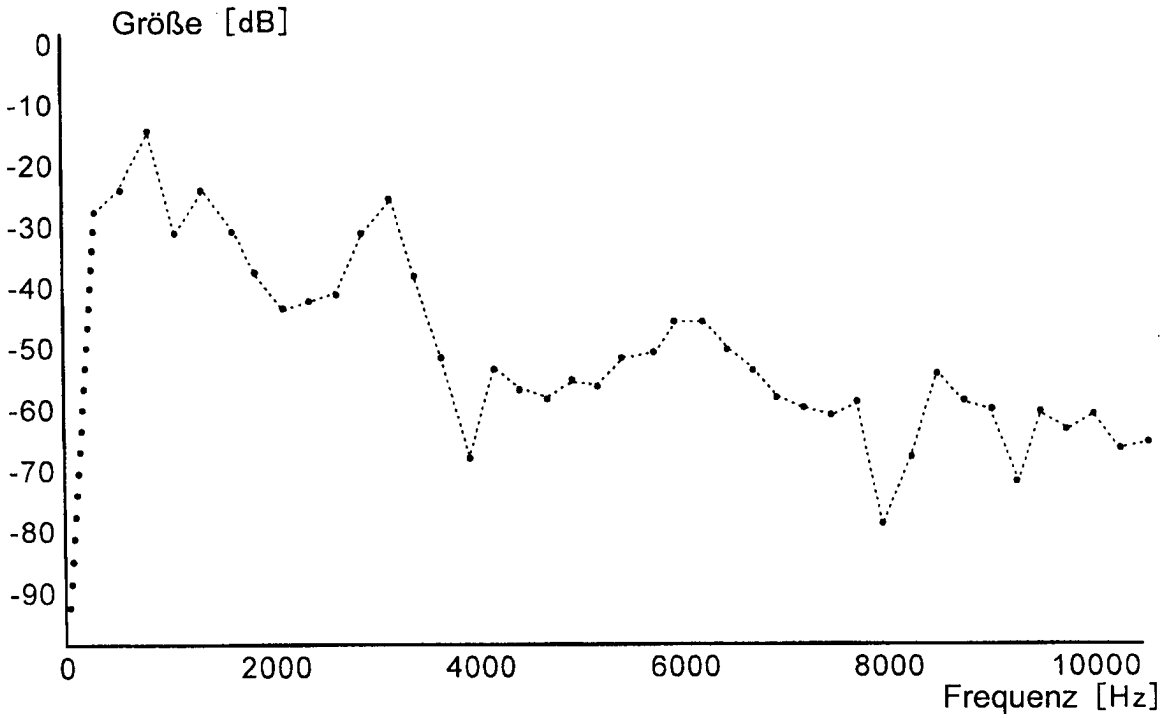


FIG. 3

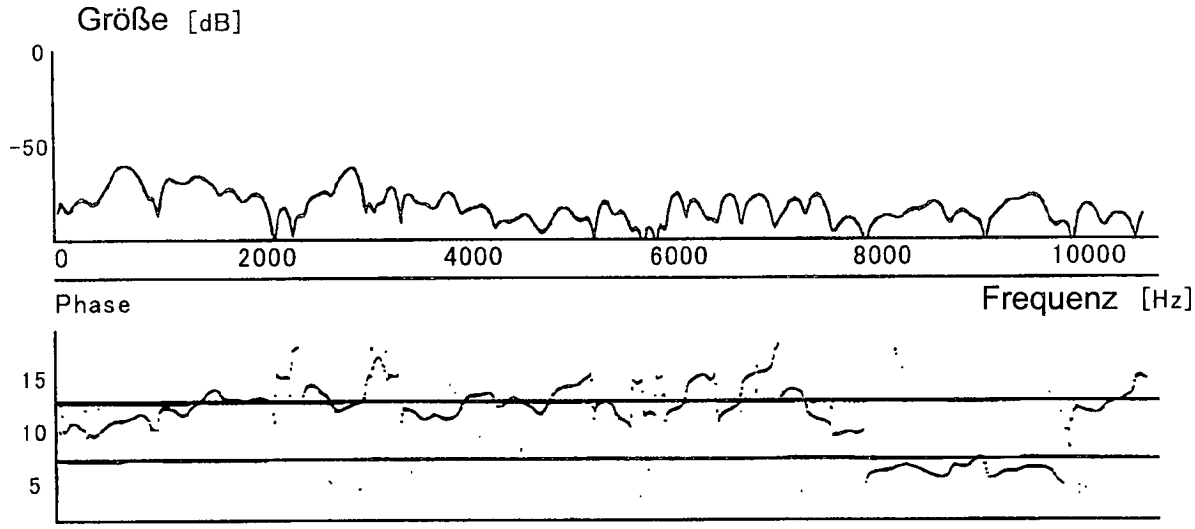

Größe [dB]
0
-50
0
2000
4000
6000
8000
10000
Frequenz [Hz]
Phase
15
10
5

FIG. 4

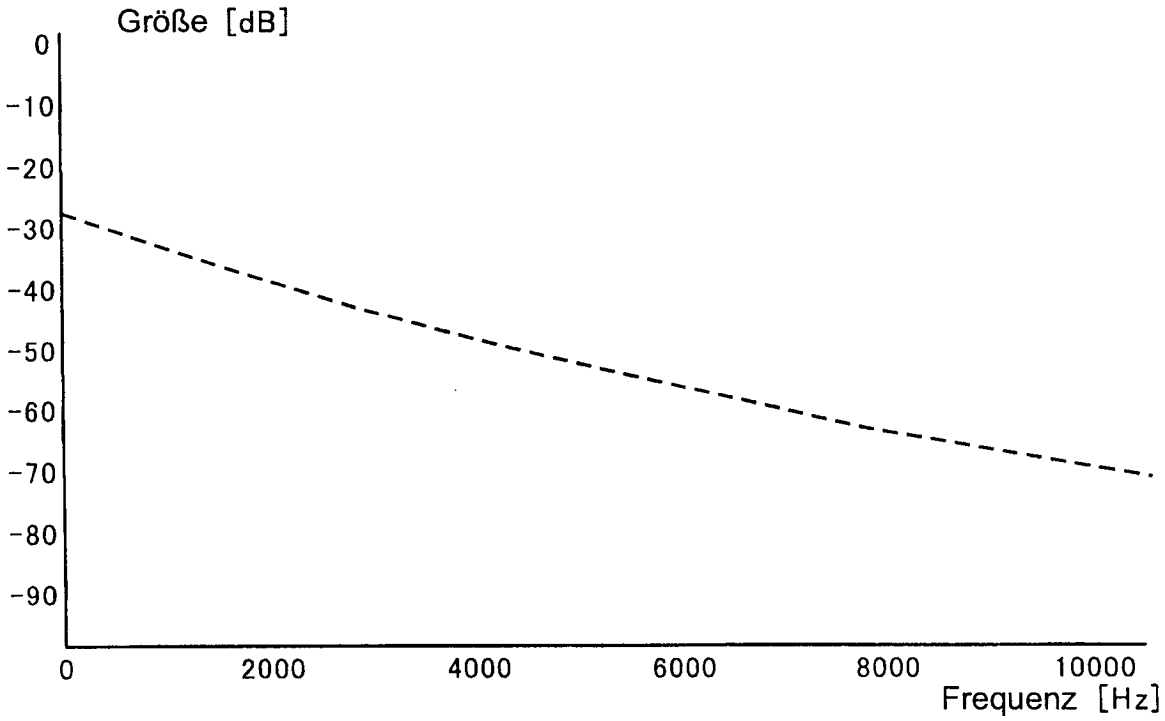

Größe [dB]
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
2000
4000
6000
8000
10000
Frequenz [Hz]

FIG. 5

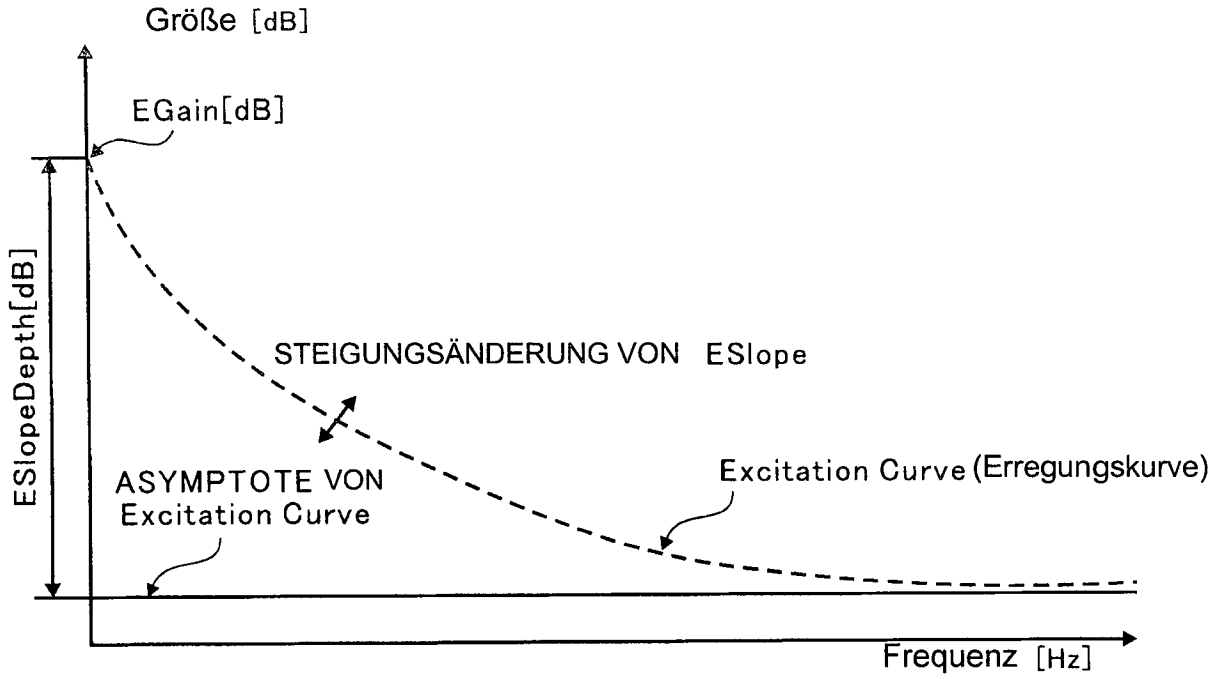

Größe [dB]
EGain[dB]
ESlopeDepth[dB]
STEIGUNGSÄNDERUNG VON ESlope
ASYMPTOTE VON
Excitation Curve
Excitation Curve (Erregungskurve)
Frequenz [Hz]

FIG. 6

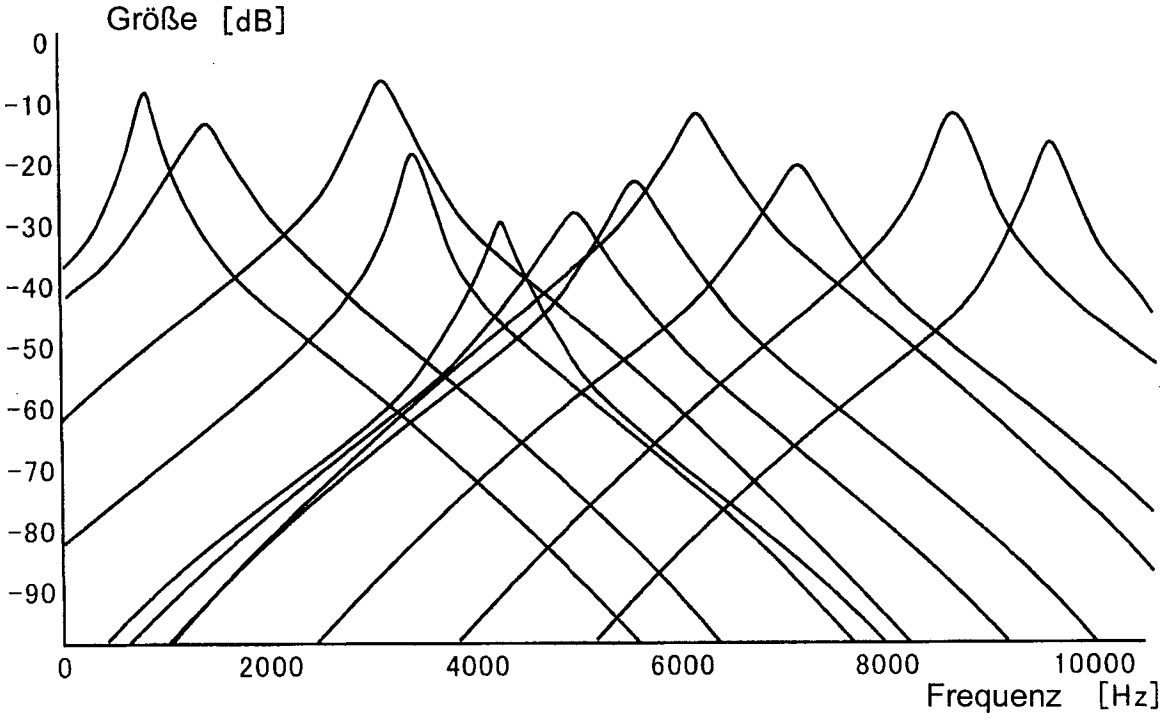

Größe [dB]
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
2000
4000
6000
8000
10000
Frequenz [Hz]

FIG. 7

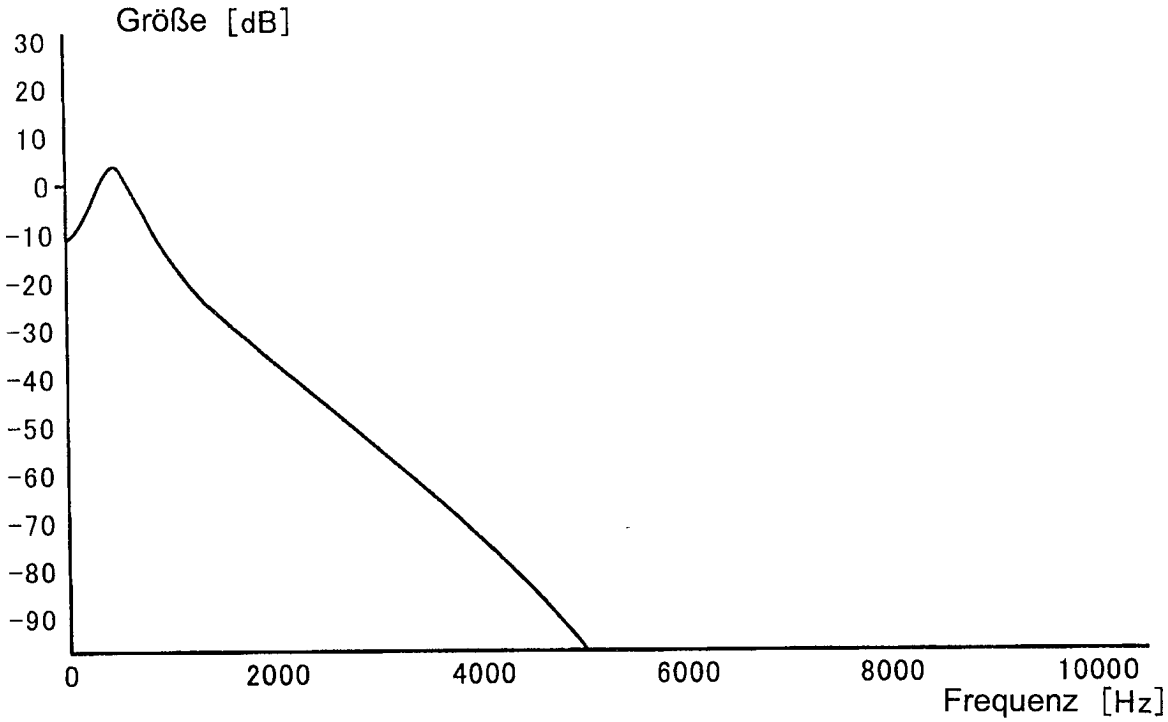

Größe [dB]
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
2000
4000
6000
8000
10000
Frequenz [Hz]

FIG. 8

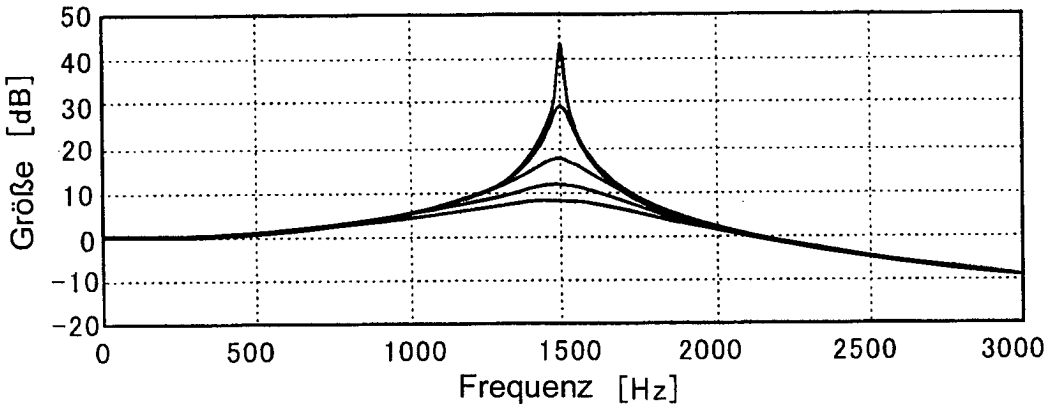

Größe [dB]
50
40
30
20
10
0
-10
-20
0
500
1000
1500
2000
2500
3000
Frequenz [Hz]

FIG. 9

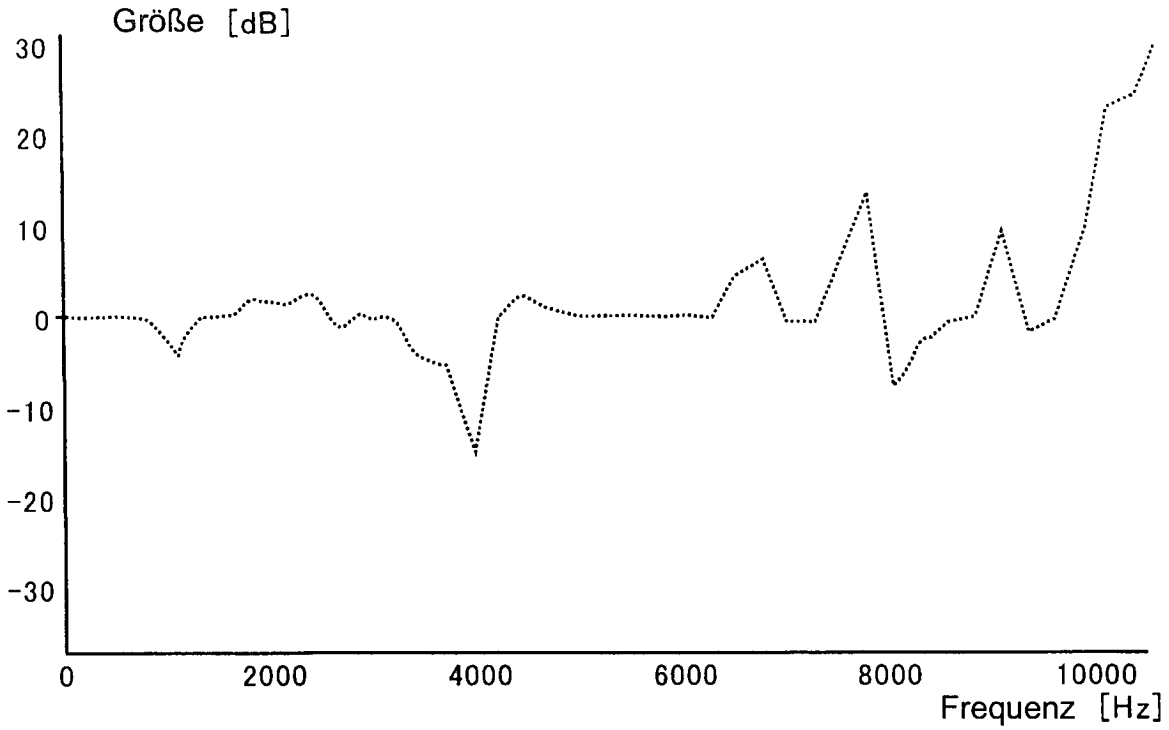

Größe [dB]
30
20
10
0
-10
-20
-30
0
2000
4000
6000
8000
10000
Frequenz [Hz]

FIG. 10

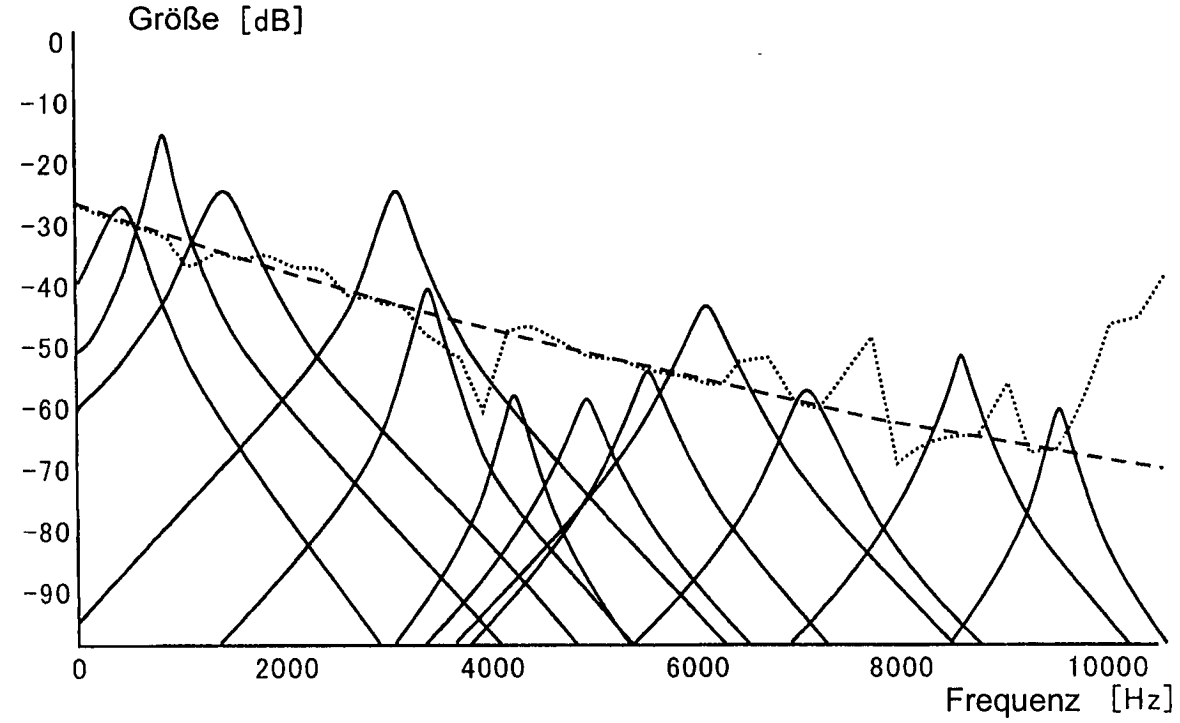

Größe [dB]
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
2000
4000
6000
8000
10000
Frequenz [Hz]

FIG. 11A

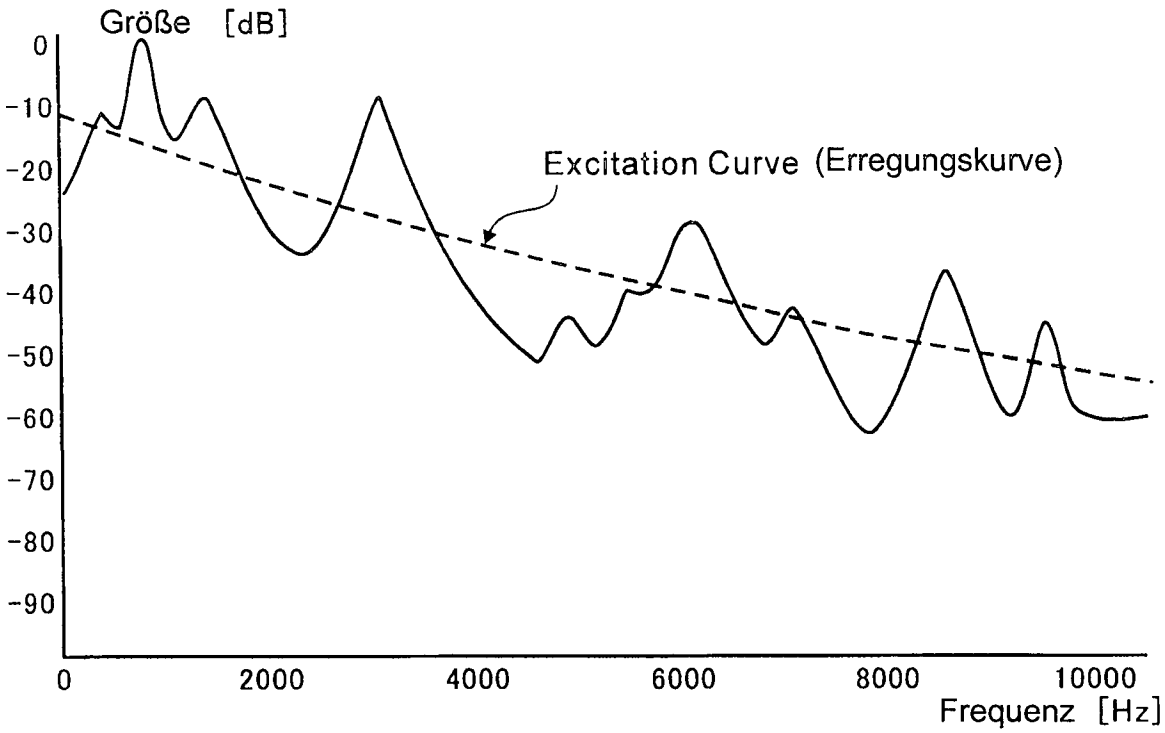


FIG. 11B

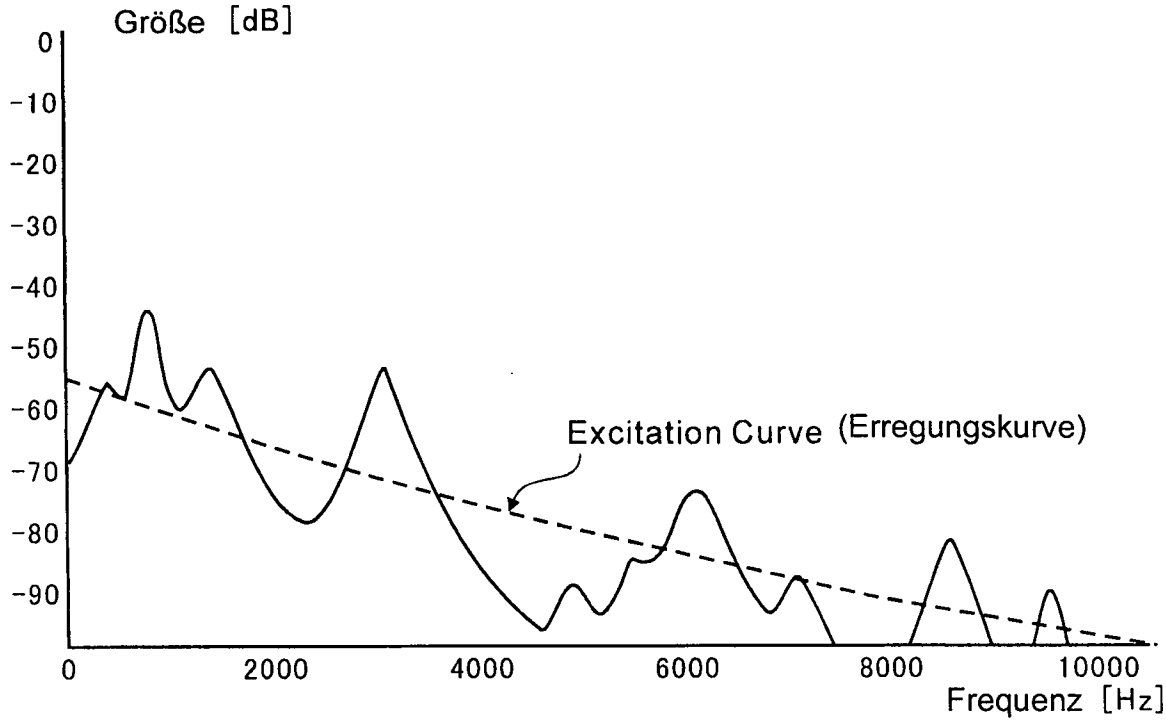


FIG. 12A

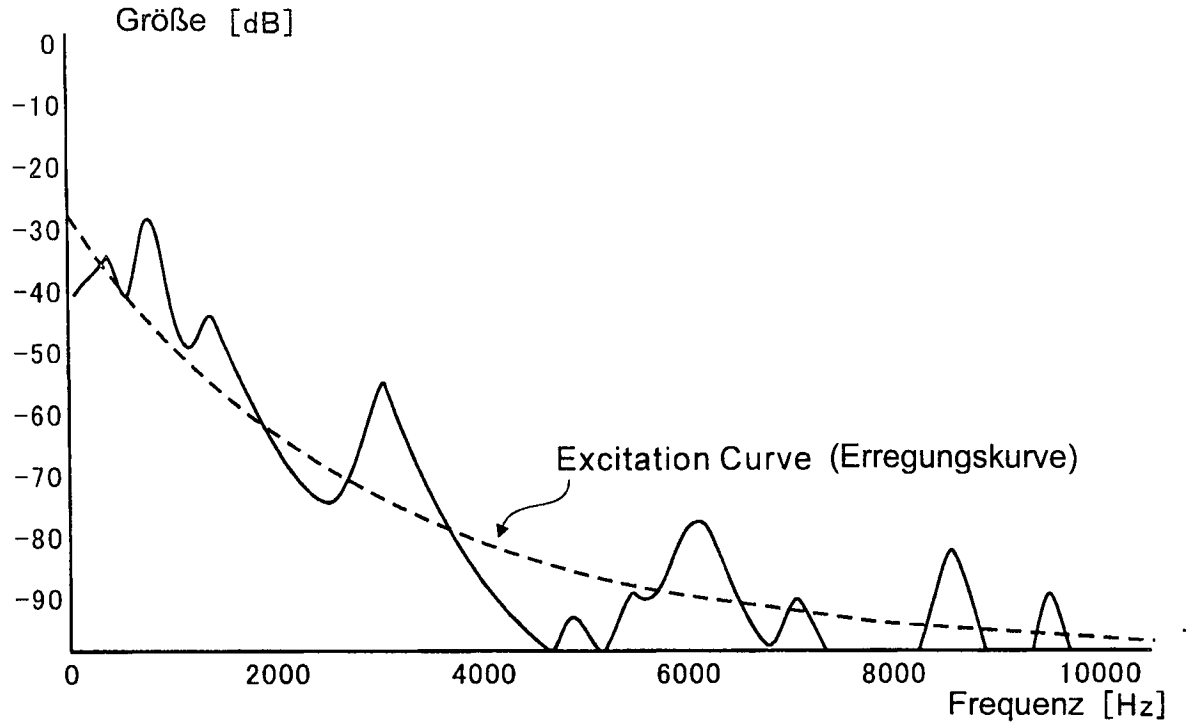


FIG. 12B

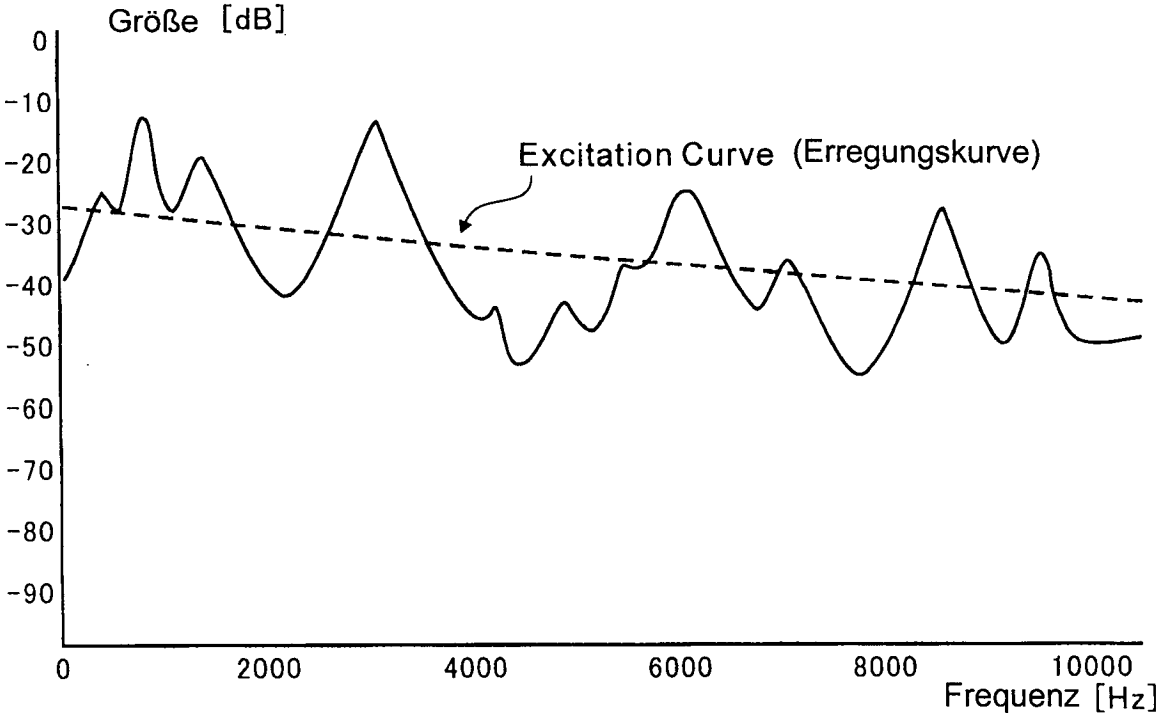


FIG. 13A

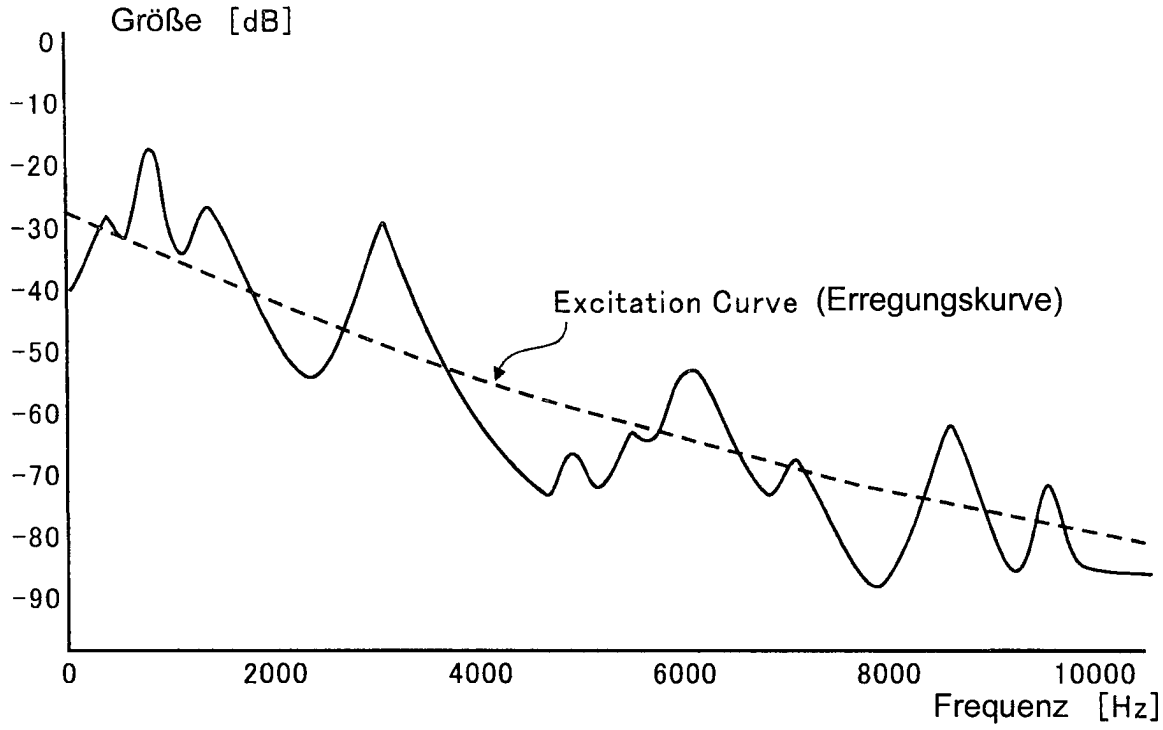


FIG. 13B

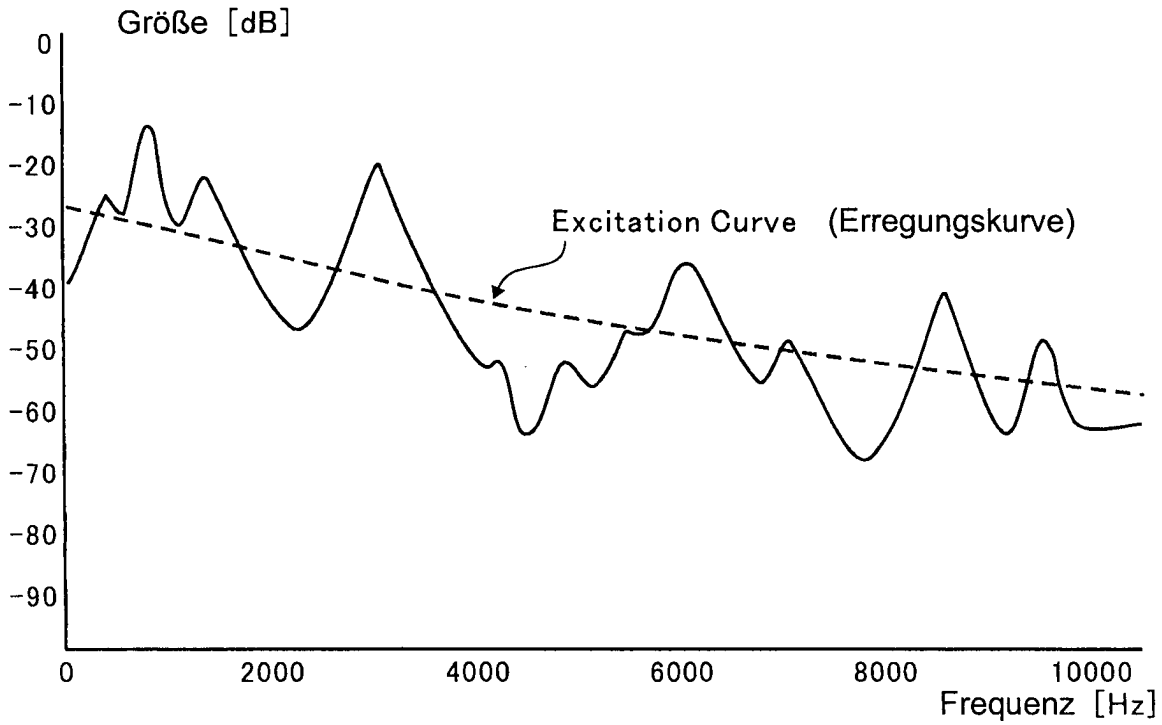


FIG. 14A

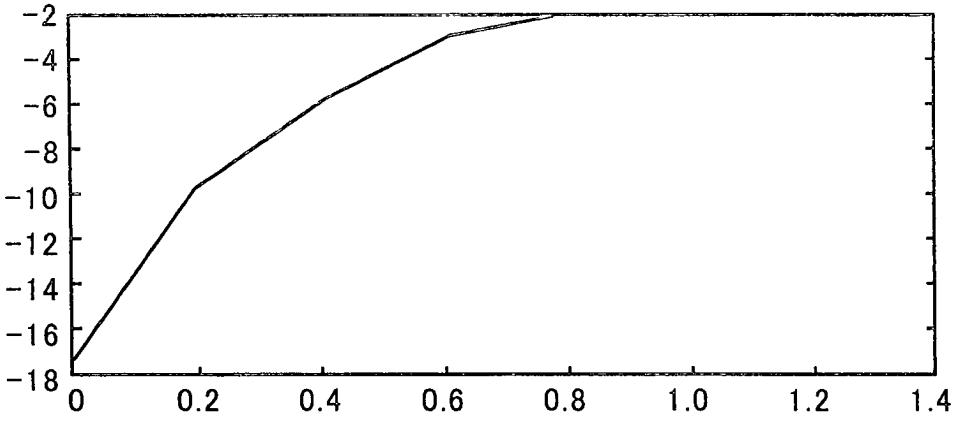


FIG. 14B

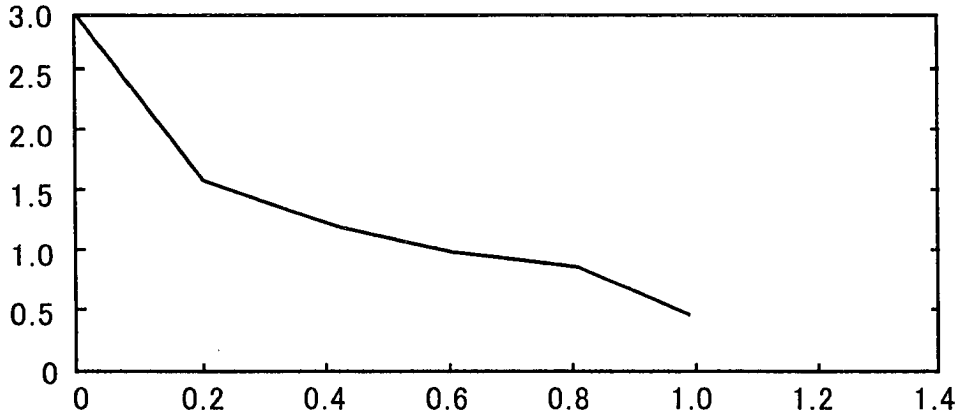


FIG. 14C

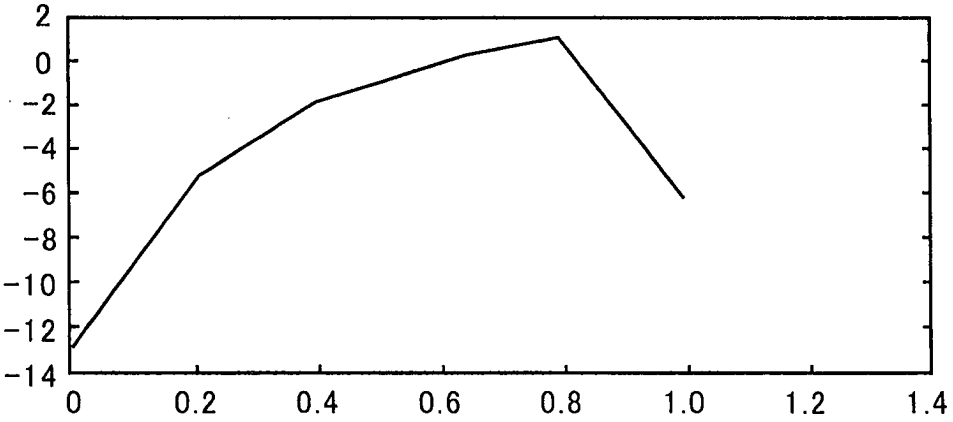


FIG. 15

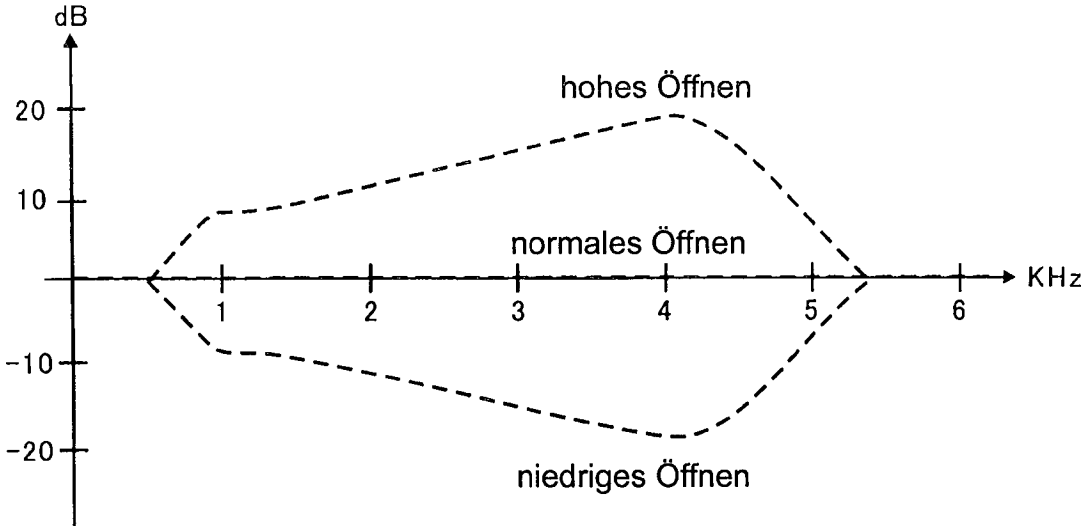
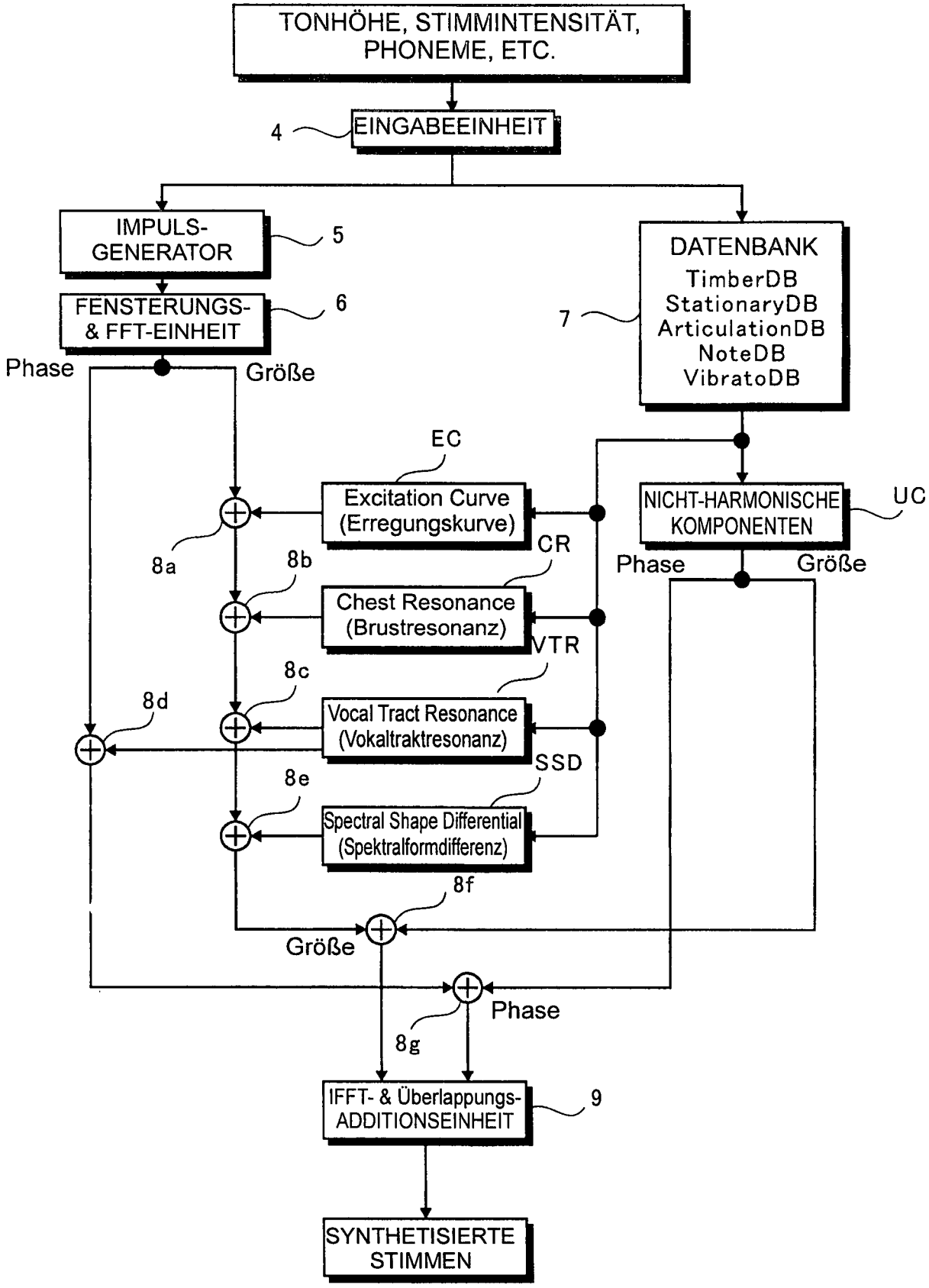

dB
20
10
-10
-20
hohes Öffnen
normales Öffnen
niedriges Öffnen
1
2
3
4
5
6
KHz

FIG. 16


TONHÖHE, STIMMINTENSITÄT, PHONEME, ETC.
4
EINGABEEINHEIT
IMPULS-GENERATOR
5
FENSTERUNGS- & FFT-EINHEIT
6
Phase
Größe
7
DATENBANK
TimberDB
StationaryDB
ArticulationDB
NoteDB
VibratoDB
EC
Excitation Curve (Erregungskurve)
8a
CR
8b
Chest Resonance (Brustresonanz)
VTR
8c
8d
Vocal Tract Resonance (Vokaltraktresonanz)
SSD
8e
Spectral Shape Differential (Spektralformdifferenz)
NICHT-HARMONISCHE KOMPONENTEN
UC
Phase
Größe
8f
Größe
Phase
8g
IFFT- & Überlappungs-ADDITIONSEINHEIT
9
SYNTHETISIERTE STIMMEN