



(11)

**EP 2 394 271 B1**

(12)

## EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:  
**22.03.2017 Patentblatt 2017/12**

(51) Int Cl.:  
**G10L 21/0364** <sup>(2013.01)</sup> *G10L 25/03* <sup>(2013.01)</sup>  
*G10L 25/18* <sup>(2013.01)</sup>

(21) Anmeldenummer: **10708882.5**

(86) Internationale Anmeldenummer:  
**PCT/AT2010/000032**

(22) Anmeldetag: **01.02.2010**

(87) Internationale Veröffentlichungsnummer:  
**WO 2010/088709 (12.08.2010 Gazette 2010/32)**

(54) **METHODE ZUR TRENNUNG VON SIGNALPFADEN UND ANWENDUNG AUF DIE VERBESSERUNG VON SPRACHE MIT ELEKTRO-LARYNX**

METHOD FOR SEPARATING SIGNAL PATHS AND USE FOR IMPROVING SPEECH USING ELECTRIC LARYNX

PROCÉDÉ DE SÉPARATION DE CHEMINEMENTS DE SIGNAUX ET APPLICATION DE CE PROCÉDÉ POUR AMÉLIORER LA QUALITÉ DE LA VOIX D'UN LARYNX ÉLECTRONIQUE

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

(30) Priorität: **04.02.2009 AT 1932009**

(43) Veröffentlichungstag der Anmeldung:  
**14.12.2011 Patentblatt 2011/50**

(73) Patentinhaber: **Heimomed Heinze GmbH & Co. KG**  
**50170 Kerpen (DE)**

(72) Erfinder:  
• **HAGMÜLLER, Martin**  
**A-8041 Graz (AT)**  
• **KUBIN, Gernot**  
**A-8046 Graz (AT)**

(74) Vertreter: **Lenzing Gerber Stute**  
**PartG von Patentanwälten m.b.B.**  
**Bahnstraße 9**  
**40212 Düsseldorf (DE)**

(56) Entgegenhaltungen:

- **COLE D ET AL: "Application of noise reduction techniques for alaryngeal speech enhancement" TENCON '97. IEEE REGION 10 ANNUAL CONFERENCE. SPEECH AND IMAGE TECHNOLOGIES FOR COMPUTING AND TELECOMMUNICATIONS., PROCEEDINGS OF IEEE BRISBANE, QLD., AUSTRALIA 2-4 DEC. 1997, NEW YORK, NY, USA, IEEE, US LNKD-DOI:10.1109/TENCON.1997.648252, Bd. 2, 2. Dezember 1997 (1997-12-02), Seiten 491-494, XP010264286 ISBN: 978-0-7803-4365-8**
- **HERMANSKY H ET AL: "RASTA processing of speech" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING USA, Bd. 2, Nr. 4, Oktober 1994 (1994-10), Seiten 578-589, XP002590490 ISSN: 1063-6676 DOI: 10.1109/89.326616**
- **NIU H-J ET AL: "ENHANCEMENT OF ELECTROLARYNX SPEECH USING ADAPTIVE NOISE CANCELLING BASED ON INDEPENDENT COMPONENT ANALYSIS" MEDICAL AND BIOLOGICAL ENGINEERING AND COMPUTING, SPRINGER, HEILDELBURG, DE LNKD-DOI:10.1007/BF02349975, Bd. 41, Nr. 6, 1. November 2003 (2003-11-01), Seiten 670-678, XP001201568 ISSN: 0140-0118**

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

**EP 2 394 271 B1**

- Takayuki Arai, Keisuke Kinoshita, Nao Hodoshima, Akkio Kusumoto, Tomoko Kitamura: "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments" Acoustical Science and Technology Bd. 23, Nr. 4, 31. Dezember 2002 (2002-12-31), Seiten 229-232, XP002590491  
Gefunden im Internet:  
URL: [http://splab.net/papers/2002/2002\\_02.pdf](http://splab.net/papers/2002/2002_02.pdf)  
[gefunden am 2010-07-05]
- DRULLMAN R ET AL: "EFFECT OF REDUCING SLOW TEMPORAL MODULATIONS ON SPEECH RECEPTION" THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, AMERICAN INSTITUTE OF PHYSICS FOR THE ACOUSTICAL SOCIETY OF AMERICA, NEW YORK, NY, US LNKD- DOI:10.1121/1.409836, Bd. 95, Nr. 5, PART 01, 1. Mai 1994 (1994-05-01), Seiten 2670-2680, XP000447919 ISSN: 0001-4966

## Beschreibung

**[0001]** Bei der Erfindung handelt es sich um ein Verfahren zur Verbesserung der Sprachqualität eines Elektro-Larynx (EL) Sprechers, wobei das Sprachsignal des Sprechers über geeignete Mittel digitalisiert wird. Unter geeigneten Mitteln werden hier beispielsweise ein Mikrofon mit zugehörigem Analog/Digital-Umsetzer, ein Telefon oder andere Methoden unter Verwendung von elektronischem Equipment verstanden.

**[0002]** Bei einem EL handelt es sich um ein Gerät zur Bildung einer künstlichen Ersatzstimme, beispielsweise für Patienten, denen operativ der Kehlkopf entfernt wurde. Der EL wird dabei an der Unterseite des Kiefers angesetzt; ein Tongenerator mit einer bestimmten Frequenz bringt die Luft in der Mundhöhle über die Weichteile an der Unterseite des Kiefers zum Vibrieren. Diese Schwingungen werden dann durch die Artikulationsorgane moduliert, so dass ein Sprechen möglich wird. Da allerdings der Tongenerator meistens nur mit einer Frequenz arbeitet, klingt die Stimme monoton und unnatürlich, bzw. "roboterhaft".

**[0003]** Weiters ist von Nachteil, dass die Vibration des EL die Wahrnehmung des Sprechens stört oder sogar übertönt, weil nur ein Teil des Schalls in der Mundhöhle artikuliert wird. Die direkt vom Gerät oder an der Übergangsstelle am Hals austretenden Anteile überlagern die artikulierten Teile und setzen die Verständlichkeit herab. Dies ist besonders bei Sprechern der Fall, die einer Strahlentherapie im Halsbereich unterzogen wurden, wodurch sich die Gewebestruktur versteift. Es wurden daher verschiedene Methoden entwickelt, die das Nutzsignal - also die artikulierten Schwingungen - gegenüber dem Störsignal - also dem Direktschall, bzw. der unmodulierten Vibration des EL - verstärken sollen.

**[0004]** Diese Methoden kommen dabei überwiegend in Situationen zum Einsatz, bei denen der Zuhörer dem abgestrahlten Schall nicht unmittelbar ausgesetzt ist, sondern elektronische Mittler verwendet werden, beispielsweise beim Telefonieren, bei Schallaufzeichnungen oder allgemein beim Sprechen über Mikrofon und Verstärker.

**[0005]** In der US 6,359,988 B1 wird ein EL-Stimmsignal einer Cepstrum-Analyse unterworfen und mit der Sprache eines Normalsprechers überlagert, wodurch sich die Tonlagenveränderung des mit EL Sprechenden natürlicher gestalten lässt; gleichzeitig wird dadurch auch der Anteil des abgestrahlten Direktschalls am Signal unterdrückt. Nachteil an dieser Lösung ist vor allem, dass zu jeder Aussage eines EL-Sprechers zeitgleich die gleiche Aussage eines gesunden (also ohne EL sprechenden) Sprechers benötigt wird, was praktisch kaum realisierbar ist.

**[0006]** Eine weitere Lösung zeigt die US 6,975,984 B2, in der eine Lösung zum Verbessern eines EL-Sprachsignals in der Telephonie beschrieben wird. Dabei wird in einem digitalen Signalprozessor das Sprachsignal derart bearbeitet, dass das brummende Grundgeräusch des EL

erkannt und aus dem Sprachsignal entfernt wird. Das Sprachsignal wird dafür in eine stimmhafte und eine stimmlose Komponente aufgeteilt und getrennt verarbeitet. Der stimmhafte Teil wird blockweise fouriertransformiert, frequenzgefiltert (Grundfrequenz und Harmonische werden weiterverwendet), rücktransformiert und in der Folge vom gesamten Originalsignal subtrahiert. Übrig bleibt der stimmlose Anteil des Originalsignals. Alternativ wird auch vorgeschlagen, den stimmhaften Anteil über Tiefpass zu filtern, im Falle der Erkennung einer Sprachpause völlig auszufiltern und den stimmlosen Anteil hinterher zu überlagern.

**[0007]** Das Dokument "Enhancement of Electrolaryngeal Speech by Adaptive Filtering" von Carol Y. Espy-Wilson et al. (JSLHR, 41: 1253-1264, 1998) beschreibt eine Methode zur Verbesserung der Sprachqualität eines EL-Sprechers. Das Grundgeräusch des EL wird dabei mittels adaptiver Filterung an das durch das EL-Grundgeräusch gestörte Sprachsignal (bzw. das zu Sprache artikuliert EL-Grundgeräusch) angeglichen; in einem weiteren Schritt werden die Signale voneinander abgezogen. Übrig bleibt ein Fehlersignal, das zur Kontrolle und Anpassung der Filterparameter mit dem Ziel der Minimierung des Fehlersignals verwendet wird. Das Fehlersignal in der vorliegenden Methode ist das vom EL-Grundgeräusch befreite Sprachsignal. Die Annahme dabei ist, dass zwar das Störsignal im Sprachsignal mit dem EL-Grundgeräusch korreliert ist, das interessierende Sprachsignal aber unabhängig von den anderen Signalen ist, dass also quasi das störende Grundgeräusch und das Sprachsignal von unterschiedlichen Quellen herrühren.

**[0008]** Das Dokument "Enhancement of Electrolarynx Speech Based on Auditory Masking" von Hanjun Liu et al. (IEEE Transactions on Biomedical Engineering, 53(5): 865-874, 2006) beschreibt einen Subtraktionsalgorithmus zur Signalverbesserung eines EL-Sprechenden, insbesondere im Bezug auf Umgebungslärm.

**[0009]** Im Gegensatz zu anderen Methoden, die fixe Subtraktionsparameter vorsehen, werden bei diesem Algorithmus die Subtraktionsparameter im Frequenzbereich adaptiert, basierend auf auditorischer Maskierung. Dabei wird davon ausgegangen, dass Sprache und Hintergrundgeräusche unkorreliert sind und deshalb der Hintergrundlärm abgeschätzt und im Frequenzbereich vom Signal abgezogen werden kann.

**[0010]** Diesen Lösungen ist gemeinsam, dass Methoden basierend auf einem Modell verwendet werden, wonach Sprache und Störsignal (also Umgebungsgeräusche, aber auch das Grundgeräusch des EL) statistisch unabhängig, bzw. unkorreliert sind.

**[0011]** Aufgrund dieser Annahme erfolgt die Implementierung der genannten Methoden auf sehr aufwändige Art und Weise. Wenn versucht wird, den Direktschall mit einem (adaptiven) Notchfilter zu unterdrücken, wird dadurch auch die Qualität des Sprachsignals vermindert, das dann wie ein Flüstern klingt; Sprachsignal und Stör-

geräusch liegen auf den gleichen Harmonischen.

**[0012]** Die US 2005/0004604 A1 beschreibt eine Larynx-Lösung, bei der ein Tongeber und ein Mikrofon direkt vor dem Mund eines Anwenders platziert werden, wobei der Tongeber einen Ton mit geringer Lautstärke abgibt und das Signal für die Weiterverarbeitung über das Mikrofon aufgenommen wird. Bei der Weiterverarbeitung wird das Signal im Wesentlichen mit einem Kammfilter gefiltert, um die Harmonischen des Signals zu reduzieren bzw. zu entfernen. Dabei wird aber die Qualität des Sprachsignals stark in Mitleidenschaft gezogen.

**[0013]** In WO 2006/099670 A1 ist eine Vorrichtung zur Überwachung der Atemwege beschrieben, wobei Schall im hörbaren Frequenzbereich in die Atemwege eines Objekts eingebracht wird und aus dem reflektierten bzw. verarbeiteten Schall der Zustand der Atemwege ermittelt wird. So ist es beispielsweise möglich, eine Verlegung der Atemwege nachzuweisen. In einer Variante der Erfindung wird mittels der FFT (Fast-Fourier-Transformation) das Überschreiten von bestimmten Schwellenwerten überprüft, woraus Rückschlüsse auf die Behandlung des gemessenen Signals gezogen werden.

**[0014]** Es ist eine Aufgabe der Erfindung, die oben genannten Nachteile des Stands der Technik zu überwinden und die Sprachqualität von EL-Anwendern bei Verwendung von elektronischen Mitteln wie beispielsweise Mikrofonen zu verbessern.

**[0015]** Diese Aufgabe wird mit einem Verfahren der eingangs erwähnten Art erfindungsgemäß durch die folgenden Schritte gelöst:

- a) Aufteilen eines einkanaligen Sprachsignals  $S(w,t)$ , das aus der Summe der Komponenten eines zeitvarianten Anteils  $x(w)H(w,t)$  und eines zeitinvarianten Anteils  $x(w)F(w)$  besteht, in eine Reihe von Frequenzkanälen durch Überführen vom Zeitbereich in einen diskreten Frequenzbereich,
- b) Herausfiltern des zeitinvarianten Anteils  $x(w)F(w)$  durch Herausfiltern der Modulationsfrequenz 0 Hz mittels eines Hochpass- bzw. Notchfilters in jedem Frequenzkanal, und
- c) Rücktransformieren des gefilterten Sprachsignals vom Frequenzbereich in den Zeitbereich und Zusammenführen zu einem einkanaligen Ausgangssignal.

**[0016]** Die Erfindung macht sich ein verbessertes Modell der Anwendung eines EL zunutze, wonach das zu einem Sprachsignal artikulierte EL-Grundgeräusch sowie die unveränderten Anteile des EL, die die Wahrnehmung des Sprachsignals stören, von einer gemeinsamen Quelle, nämlich dem EL, kommen. Da das störende unartikulierte Grundgeräusch des EL im Modulationsbereich als zeitlich invariantes Signal erkennbar ist, lässt es sich durch geeignetes Vorgehen leicht ausfiltern. Es erfolgt also eine Trennung nicht von Signalquellen, sondern von Ausbreitungswegen (eines Ausbreitungsweges

durch die Artikulationsorgane eines Sprechers, ein weiterer Ausbreitungsweg von der Anwendungsstelle am Hals des Sprechers direkt zum Ohr des Zuhörers, bzw. zum Mikrofon oder Aufnahmemittel).

**[0017]** Dem Fachmann ist eine Vielzahl von Möglichkeiten bekannt, ein digitalisiertes, einkanaliges Signal in den Frequenzbereich zu überführen und so in eine Reihe von Frequenzkanälen aufzuteilen. In jedem Frequenzkanal wird die Modulationsfrequenz des EL durch geeignete Filter - z.B. Notch- oder Hochpassfilter, angewandt auf den Betrag - unterdrückt und so die Qualität der artikulierten Signalanteile verbessert.

**[0018]** Ähnliche Verfahren aus dem Stand der Technik betrachten die artikulierten Anteile sowie die unveränderten Anteile als von verschiedenen Quellen kommend und wählen diesem Modell entsprechende Herangehensweisen, beispielsweise Filterung mittels Bandpassfiltern, die dann allerdings auch das Sprachsignal dämpfen.

**[0019]** Das erfindungsgemäße Verfahren zielt also darauf ab, die Verständlichkeit der Sprache von EL-Anwendern zu erhöhen bzw. das Signal angenehmer und "menschlicher" zu machen. Ziel ist es, den Direktschall aus dem EL bei Kommunikation über elektronische Mittel (z.B. Telefon) zu reduzieren bzw. zu eliminieren.

**[0020]** Die Realisierung des erfindungsgemäßen Verfahrens kann beispielsweise durch ein Software-Plugin, als fest verdrahtete Lösung oder auch als Analogschaltung erfolgen.

**[0021]** Aus der Vielzahl bekannter Methoden zur Überführung eines Signals in den Frequenzbereich bzw. zurück erfolgt die Überführung in Schritt a) des erfindungsgemäßen Verfahrens günstigerweise mittels Fourier-Transformation und die Rücktransformation in Schritt c) mittels inverser Fourier-Transformation. Die Überführung erfolgt blockweise (z.B. Blöcke von 20 ms) in kurzen Abständen (Auffrischung beispielsweise alle 10 ms). Die Aufteilung des Signals in eine Reihe von Frequenzkanälen erfolgt beim Überführen des Signals in den Frequenzbereich.

**[0022]** In einer Variante der Erfindung erfolgt die Überführung des Sprachsignals in Schritt a) und die Rücktransformation in Schritt c) mit einer entsprechenden Filterbank.

**[0023]** Die Ergebnisse des erfindungsgemäßen Verfahrens lassen sich weiter verbessern, wenn vor der Filterung in Schritt b) eine Signal-Kompression erfolgt und nach Schritt b) eine Dekompression erfolgt. Durch die Kompression kann verhindert werden, dass bei hohen Amplituden deren Änderungen derart dominant sind, dass die Änderungen kleiner Amplituden nicht berücksichtigt werden. Durch die Kompression werden also relative Änderungen für das Filter besser sichtbar.

**[0024]** In einer weiteren Ausführung des erfindungsgemäßen Verfahrens erfolgt vor der Rücktransformation in Schritt c) eine Gleichrichtung der negativen Signalanteile.

**[0025]** Im Folgenden wird die Erfindung anhand eines

nicht einschränkenden Ausführungsbeispiels, das in der Zeichnung dargestellt ist, näher erläutert. In dieser zeigt schematisch:

Fig. 1 eine vereinfachte Darstellung der Verwendung eines EL und die auftretenden Signalfade,

Fig. 2 eine vereinfachte Darstellung der Situation, in der die erfindungsgemäße Methode Anwendung findet und

Fig. 3 ein Blockschaltbild der erfindungsgemäßen Methode.

**[0026]** In Fig. 1 sind die verschiedenen Übertragungswege des Signals eines EL 1 skizziert. Dabei ist am Hals eines Sprechers 2 ein EL 1 angeordnet. Der vom EL 1 erzeugte Schall breitet sich einerseits durch die normalen Sprachkanäle (Mund und Nase) 5 des ersten Sprechers 2 aus und wird dort zu Sprache artikuliert; dieses erste Signal 3 ist deutlich veränderlich, bzw. zeit-variant. Am Ohr eines Zuhörers 4 kommt neben diesem zeit-varianten Signal 3 auch ein zweites Signal 6 (in Fig. 1 strichpunktiert dargestellt) in Form des Direktschalls des EL 1 an, wobei dieses Signal 4 weitgehend stationär ist und daher als zeitlich invariant angenommen wird. Der zweite Teil 6 des Gesamtsignals, also das Grundgeräusch des EL 1, wird vom Zuhörer 4 als Störsignal wahrgenommen und verringert die Verständlichkeit der Sprache des Sprechers 2. Die ursprüngliche Anregung mittels des EL 1 wird also über zwei verschiedene Pfade übertragen.

**[0027]** Zwar bezieht sich die Erfindung auf die Verbesserung der Sprachqualität eines EL-Sprechers bei Verwendung von elektronischen Mittlern - anstatt eines Zuhörers würden die Signale also beispielsweise mit einem Mikrophon aufgenommen werden. Zur Illustration der Ausgangslage wurde allerdings aus Gründen der Verständlichkeit dieses allgemeine Modell gewählt.

**[0028]** Fig. 2 zeigt eine vereinfachte Modelldarstellung der Situation, auf die die erfindungsgemäße Methode zur Unterdrückung eines störenden zweiten Signals 6 (siehe Fig. 1) angewendet wird. Es ist gut erkennbar, dass es bei der erfindungsgemäßen Methode nicht zu einer Trennung von Signalquellen, sondern von Ausbreitungswegen kommt.

**[0029]** Ein Quellensignal  $x(w)$  von einer Signalquelle 7 breitet sich über zwei verschiedene Signalfade aus. Im ersten Signalfade wird das Ausgangssignal durch ein zeitvariantes Filter  $H(w, t)$  zu einem zeitvarianten Signal  $x(w)H(w, t)$  moduliert. Im zweiten Signalfade wird das Ausgangssignal nur durch ein zeitinvariantes Filter  $F(w)$  zu einem Signal  $x(w)F(w)$  verändert.

**[0030]** Die Signale der beiden Pfade werden dann in einem Empfänger 8 - z.B. dem Ohr eines Zuhörers, einem Mikrophon o.ä. - zu einem zur Messung zur Verfügung stehenden Signal  $S(w, t)$  summiert. Das Signal besteht dann aus der Summe der Komponenten,

$$S(w, t) = x(w)H(w, t) + x(w)F(w).$$

**[0031]** Es können nun die Signalteile vom zeitinvarianten und vom zeitvarianten Signalfade getrennt werden, indem entweder alle Signalanteile, die sich zeitlich ändern, bzw. zeitlich konstant sind, gedämpft werden. Man erhält also beispielsweise als Ergebnis nur den zeitvarianten Anteil  $S_1(w, t) \sim x(w)H(w, t)$ .

**[0032]** Bei der Anwendung für Sprache mit EL überlagert der unartikulierte Signalanteil  $x(w)F(w)$  (also das Grundgeräusch des EL) das zeitvariante Sprachsignal  $x(w)H(w, t)$  und bewirkt dadurch einen Verständlichkeitsverlust für das Sprachsignal. Die Sprachverständlichkeit wird verbessert, indem der zeitvariante Signalanteil vom zeitinvarianten Signalanteil getrennt wird.

**[0033]** Fig. 3 zeigt eine mögliche Umsetzung der erfindungsgemäßen Methode. Dabei kann am Eingang ein beliebiges digitales Sprachsignal 9 von einem Sprecher mit EL anliegen. In einem ersten Schritt 10 wird unter Anwendung der Kurzzeit-Fouriertransformation das Sprachsignal 9 blockweise in den Frequenzbereich transformiert und so in eine Reihe von Frequenzkanälen aufgeteilt. Der Fachmann kann hier aus verschiedenen etablierten Methoden zur Transformation eines Signals vom Zeit- in den Frequenzbereich wählen; neben der Fourier-Transformation findet beispielsweise auch die Diskrete Kosinustransformation Anwendung - Voraussetzung für eine erfindungsgemäße Anwendung ist allerdings, dass die Transformation umkehrbar ist. Das Signal wird mit einer bestimmten Auffrischungsrate (z.B. 10ms) in Blöcke von beispielsweise 20 ms Länge aufgeteilt, die jeweils in eine Reihe von Frequenzkanälen 11 aufgefächert werden. Das ursprünglich einkanalige Sprachsignal 9 wird also in eine Vielzahl von Frequenzbereichen aufgespaltet, die sich als Folge der Zeit ändern. Das Frequenzsignal ist komplex, es wird aber in weiterer Folge nur der Absolutbetrag modifiziert, die Phase 15 bleibt unverändert.

**[0034]** In Schritt 10 kann auch eine Filterbank verwendet werden, wobei die Abtastrate des Signals nach der Filterbank reduziert wird. Das Reduzieren der Abtastrate entspricht dabei der Blockbildung bei Anwendung der Fourier-Transformation.

**[0035]** In einem weiteren Funktionsblock 12 wird nun jeder Frequenzkanal 11 gefiltert, beispielsweise mit einem Hochpass- bzw. Notchfilter. Diese Filterung erlaubt das Ausfiltern bestimmter Frequenzen - in der Tontechnik werden mit Notchfiltern schmalbandige Störungen beseitigt. Da der EL auf einer bestimmten Frequenz oszilliert - beispielsweise 100 Hz - ergibt das Störsignal, das nicht durch die Artikulationsorgane eines Sprechers verändert ist, im Frequenzbereich Amplituden im 100 Hz-Kanal mit der Modulationsfrequenz 0 Hz - d.h., dass sich die Amplitude des EL-Signals nicht ändert. Das Störsignal ist dadurch gekennzeichnet, dass es perfekt zeitlich invariant ist. Zur Filterung des Grundgeräuschs des EL

werden ein Notch- bzw. ein Hochpassfilter verwendet. Als Grenzfrequenz für das Hochpassfilter dient dabei die Modulationsfrequenz des EL; das Notchfilter wird so gewählt, dass es genau bei der Modulationsfrequenz des EL sperrt.

**[0036]** In der realen Umsetzung wird natürlich eine perfekte zeitliche Invarianz aufgrund von Reflexionen, Brechungen, Umgebungsgeräuschen und baulicher Notwendigkeiten des EL nicht erreichbar sein. Da allerdings auch das Filter nicht auf nur eine Frequenz eingeschränkt ist, sondern einen bestimmten Frequenzbereich - in diesem Fall einen Modulationsfrequenzbereich - abdeckt, ist die Funktion der erfindungsgemäßen Methode sichergestellt.

**[0037]** In einem abschließenden Funktionsblock 13 erfolgt die Rückführung der Signale in den Zeitbereich, beispielsweise mittels inverser Fourier-Transformation und die Zusammenführung der Frequenzkanäle 11 zurück in einen Kanal mittels overlap-add. Das overlap-add Verfahren ist dabei ein dem Fachmann bekanntes Verfahren aus der digitalen Signalverarbeitung. Ergebnis ist ein einkanaliges Ausgangssignal 14, in dem das Störsignal des EL ausgefiltert oder zumindest gedämpft ist. Das Ausgangssignal kann dann weiter verarbeitet werden.

**[0038]** Bei Anwendung einer Filterbank in Schritt 10 wird die Abtastrate des Signals nach der Filterung in Schritt 12 wieder erhöht und dann wie geschildert weiterbehandelt.

**[0039]** Grundsätzlich stellen diese Ausführungen nur die wichtigsten Bestandteile der erfindungsgemäßen Methode dar; vor der Filterung im Block 12 kann das Signal komprimiert werden, nach der Filterung kann eine Dekomprimierung vorgesehen sein. Auch eine Gleichrichtung vor der Rücktransformation in den Zeitbereich kann günstig sein, da bei der Bearbeitung unerlaubte negative Werte entstehen können.

**[0040]** Die Erfindung kann beispielsweise als Zusatzgerät zum Telefonieren verwendet werden. Bei einem herkömmlichen analogen Telefon wird das Gerät einfach in den Hörer integriert. Bei einem Telefon mit integriertem Digitalem Signal Prozessor ist die Integration der Erfindung durch ein Software-Plugin möglich. Auch die Realisierung im Rahmen einer fest verdrahteten Lösung, z.B. auch in einer Analogschaltung, ist möglich.

**[0041]** Die erfindungsgemäße Methode ist auch bei Verwendung eines EL einsetzbar, bei dem zwischen zwei oder mehr Frequenzen hin- und hergeschaltet werden kann um der Sprache einen realistischeren Klang zu geben. Das gilt sowohl für diskrete Frequenzsprünge als auch für kontinuierliche Änderungen der Grundfrequenz unter der Annahme, dass die Frequenzen, zwischen denen gewechselt wird, innerhalb eines Frequenzbandes liegen, in das das Grundsignal aufgeteilt wird.

**[0042]** Die Breite des Modulationsfrequenzfilters bestimmt dabei, wie schnell sich die Frequenz ändern darf. Bei sehr langsamen, kontinuierlichen Änderungen kann sich die Frequenz bei funktionierender Unterdrückung

über den gesamten Bereich des Frequenzbandes ändern - ausschlaggebend ist nicht die Größe, sondern die Geschwindigkeit der Änderung. Beim Ein- und Ausschalten des EL, das einer schnellen Änderungen entspricht, greift die Unterdrückung erst nach einigen Millisekunden - abhängig davon, wie breit das Notchfilter gewählt ist bzw. wo die Grundfrequenz des Hochpassfilters liegt.

**[0043]** Dabei dürfen allerdings die Änderungen der Grundfrequenz nicht zu groß sein. Um die erfindungsgemäße Funktion sicher zu stellen, müssten beispielsweise die Frequenzkanäle, in die das Signal aufgeteilt wird, erweitert werden, bzw. die Filterung mittels Hochpassfilter müsste an einer etwas höheren Frequenz ansetzen.

## Patentansprüche

1. Verfahren zur Verbesserung der Sprachqualität eines Elektro-Larynx (EL) Sprechers, dessen Sprachsignal  $S(w,t)$  über geeignete Mittel digitalisiert wird, **gekennzeichnet durch** die folgenden Schritte:

- Aufteilen eines einkanaligen Sprachsignals  $S(w,t)$ , das aus der Summe der Komponenten eines zeitvarianten Anteils  $x(w)H(w,t)$  und eines zeitinvarianten Anteils  $x(w)F(w)$  besteht, in eine Reihe von Frequenzkanälen **durch** Überführen vom Zeitbereich in einen diskreten Frequenzbereich,
- Herausfiltern des zeitinvarianten Anteils  $x(w)F(w)$  **durch** Herausfiltern der Modulationsfrequenz 0 Hz mittels eines Hochpass- bzw. Notchfilters in jedem Frequenzkanal und
- Rücktransformieren des gefilterten Sprachsignals vom Frequenzbereich in den Zeitbereich und Zusammenführen zu einem einkanaligen Ausgangssignal.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die Überführung des Sprachsignals in Schritt a) mittels Fourier-Transformation und die Rücktransformation in Schritt c) mittels inverser Fourier-Transformation erfolgt.

3. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die Überführung des Sprachsignals in Schritt a) und die Synthese der Frequenzkanäle in Schritt c) mit einer Filterbank erfolgt.

4. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** vor der Filterung in Schritt b) eine Signal-Kompression erfolgt und nach Schritt b) eine Dekompression erfolgt.

5. Verfahren nach einem der Ansprüche 1 bis 4, **dadurch gekennzeichnet, dass** vor der Rücktransformation in Schritt c) eine Gleichrichtung der negativen Signalkomponenten erfolgt.

## Claims

1. Method for improving the speech quality of an electric larynx (EL) speaker, whose speech signal  $S(w,t)$  is digitised by suitable means, **characterised by** the following steps:
  - a) dividing a single-channel speech signal  $S(w,t)$ , which consists of the sum of the compounds of a time variant signal  $x(w)H(w,t)$  and a time-invariant signal  $x(w)F(w)$ , into a series of frequency channels by transforming it from a time domain into a discrete frequency domain,
  - b) filtering out the time-invariant signal  $x(w)F(w)$  by filtering out the modulation frequency 0 Hz by means of a high-pass or notch filter in each frequency channel, and
  - c) back-transforming the filtered speech signal from the frequency domain into the time domain and combining it into a single-channel output signal.
2. Method according to claim 1, **characterised in that** the transforming of the speech signal in step a) is carried out by means of a Fourier transformation and the back-transformation in step c) is carried out by means of an inverse Fourier transformation.
3. Method according to claim 1, **characterised in that** the transformation of the speech signal in step a) and the synthesis of the frequency channels in step c) is carried out with a corresponding filter bank.
4. Method according to one of claims 1 to 3, **characterised in that** before the filtering in step b) a signal compression is carried out, and after step b) a decompression is carried out.
5. Method according to one of claims 1 to 4, **characterised in that** before the back-transformation in step c) a rectification of the negative signal components is carried out.

b) filtrage de la fraction invariable dans le temps  $x(w)F(w)$  par filtrage de la fréquence de modulation 0 Hz au moyen d'un filtre passe-haut ou filtre coupe-bande à bande étroite dit filtre 'notch' dans chaque canal de fréquence, et  
c) transformation retour du signal vocal filtré, du domaine de fréquences au domaine temporel, et réunification en un signal de sortie mono-canal.

2. Procédé selon la revendication 1, **caractérisé en ce que** la conversion du signal vocal dans l'étape a) s'effectue par une transformation de Fourier, et la transformation retour dans l'étape c) s'effectue par une transformation de Fourier inverse.
3. Procédé selon la revendication 1, **caractérisé en ce que** la conversion du signal vocal dans l'étape a) et la synthèse des canaux de fréquences dans l'étape c), s'effectue à l'aide d'une banque de filtres.
4. Procédé selon l'une des revendications 1 à 3, **caractérisé en ce qu'**une compression de signal est effectuée avant le filtrage dans l'étape b), et une décompression est effectuée après l'étape b).
5. Procédé selon l'une des revendications 1 à 4, **caractérisé en ce qu'**un redressement des composantes de signal négatives est effectué avant la transformation retour dans l'étape c).

## Revendications

1. Procédé d'amélioration de la qualité vocale d'un appareil phonatoire du type électro-larynx (EL), dont le signal vocal  $S(w,t)$  est numérisé par des moyens appropriés, **caractérisé par** les étapes suivantes :
  - a) subdivision d'un signal vocal  $S(w,t)$  mono-canal, qui est constitué de la somme des composantes d'une fraction variable dans le temps  $x(w)H(w,t)$ , et d'une fraction invariable dans le temps  $x(w)F(w)$ , en une série de canaux de fréquences, par la conversion du domaine temporel à un domaine discret de fréquences,

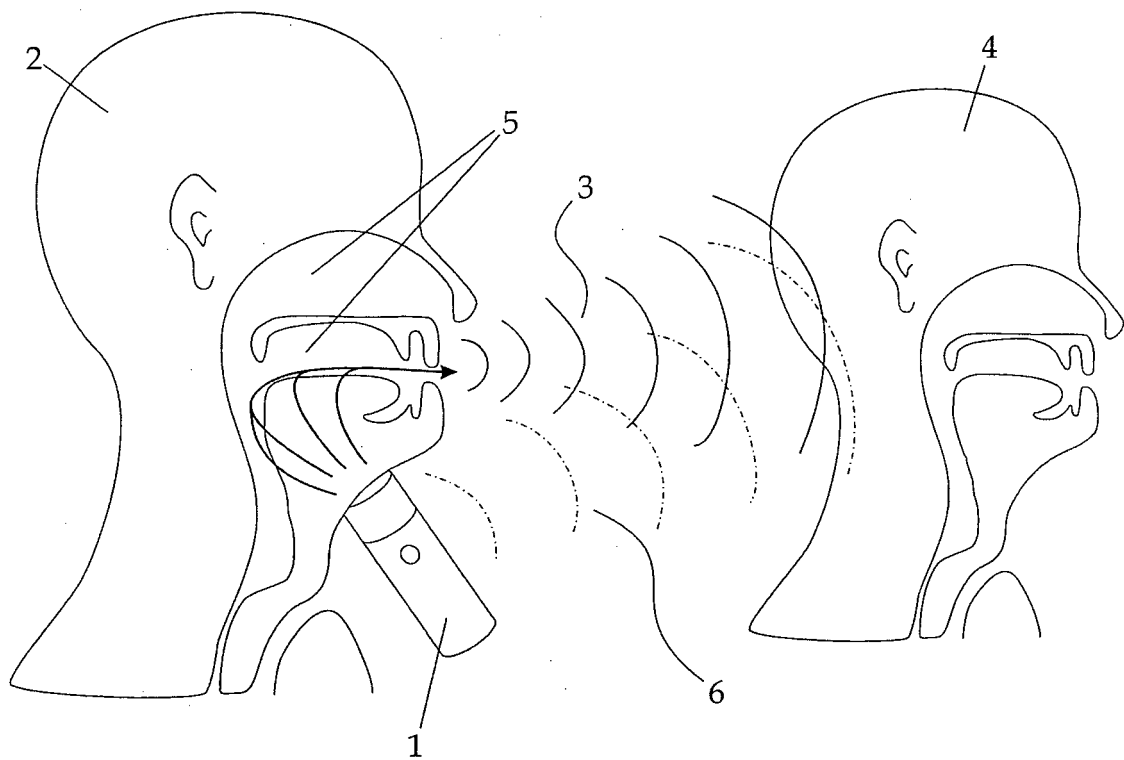


Fig. 1

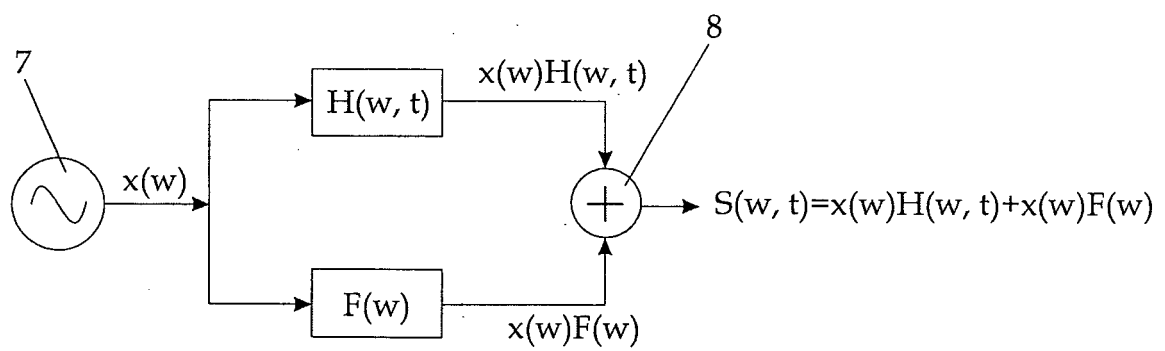


Fig. 2



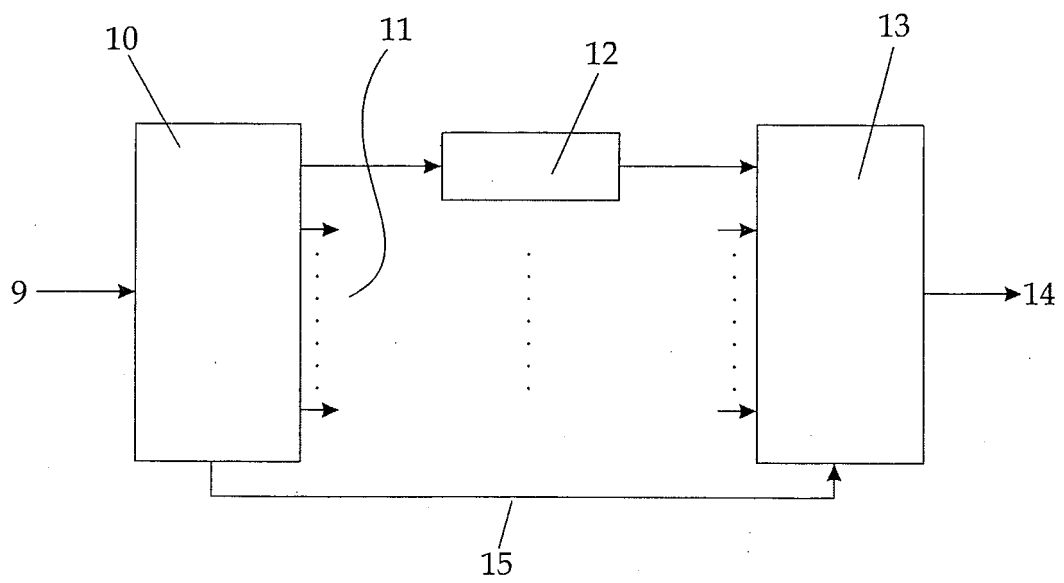


Fig. 3

## IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

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

### In der Beschreibung aufgeführte Patentdokumente

- US 6359988 B1 [0005]
- US 6975984 B2 [0006]
- US 20050004604 A1 [0012]
- WO 2006099670 A1 [0013]

### In der Beschreibung aufgeführte Nicht-Patentliteratur

- **CAROL Y. ESPY-WILSON et al.** Enhancement of Electrolaryngeal Speech by Adaptive Filtering. *JSL-HR*, 1998, vol. 41, 1253-1264 [0007]
- **HANJUN LIU et al.** Enhancement of Electrolarynx Speech Based on Auditory Masking. *IEEE Transactions on Biomedical Engineering*, 2006, vol. 53 (5), 865-874 [0008]