

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

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
12. August 2010 (12.08.2010)

(10) Internationale Veröffentlichungsnummer
WO 2010/088709 A1

(51) Internationale Patentklassifikation:
G10L 21/02 (2006.01) *A61F 2/20* (2006.01)

(21) Internationales Aktenzeichen: PCT/AT2010/000032

(22) Internationales Anmeldedatum:
1. Februar 2010 (01.02.2010)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
A 193/2009 4. Februar 2009 (04.02.2009) AT

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): **TECHNISCHE UNIVERSITÄT GRAZ** [AT/AT]; Rechbauerstraße 12, A-8010 Graz (AT). **FORSCHUNGS HOLDING TU GRAZ GMBH** [AT/AT]; Rechbauerstraße 12, A-8010 Graz (AT).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): **HAGMÜLLER, Martin** [AT/AT]; Messendorfer Straße 62, A-8041 Graz (AT). **KUBIN, Gernot** [AT/AT]; Franz-Schmidt-Weg 14, A-8046 Graz (AT).

(74) Anwalt: **PATENTANWALT SKANZLEI MATSCHNIG & FORSTHUBER**; Siebensterngasse 54, A-1071 Wien (AT).

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR SEPARATING SIGNAL PATHS AND USE FOR IMPROVING SPEECH USING ELECTRIC LARYNX

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

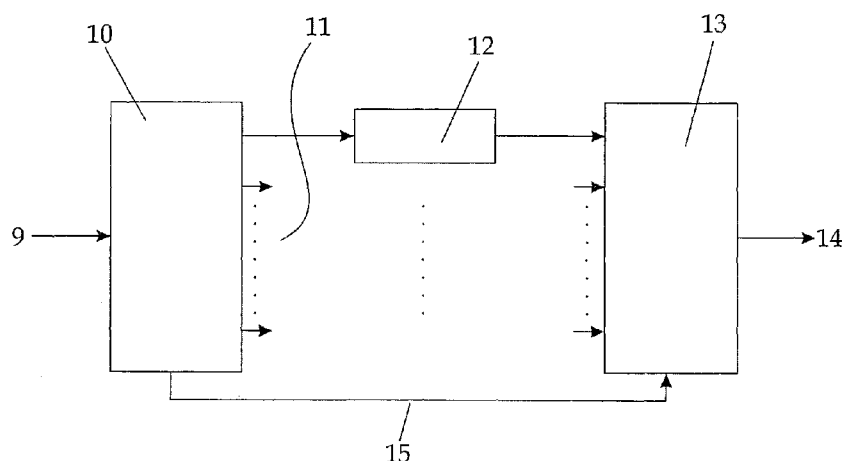


Fig. 3

(57) Abstract: In order to improve the speech quality of an electric larynx (EL) speaker, the speech signal of which is digitized by suitable means, the following steps are carried out: a) dividing a single-channel speech signal into a series of frequency channels by transferring it from a time domain into a discrete frequency domain; b) filtering out the modulation frequency of the EL by way 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.

(57) Zusammenfassung: Zur Verbesserung der Sprachqualität eines Elektro-Larynx (EL) Sprechers, dessen Sprachsignal über geeignete Mittel

[Fortsetzung auf der nächsten Seite]



WO 2010/088709 A1



Veröffentlicht:

— mit internationalem Recherchenbericht (Artikel 21 Absatz 3)

— vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eingehen (Regel 48 Absatz 2 Buchstabe h)

digitalisiert wird, werden die folgenden Schritte durchgeführt: a) Aufteilen eines einkanaligen Sprachsignals in eine Reihe von Frequenzkanälen durch Überführen vom Zeitbereich in einen diskreten Frequenzbereich, b) Herausfiltern der Modulationsfrequenz des EL 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.

METHODE ZUR TRENNUNG VON SIGNALPFADEN UND ANWENDUNG AUF DIE VERBESSERUNG VON SPRACHE MIT ELEKTRO-LARYNX

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.

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“.

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.

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.

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.

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.

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 artikulierte 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.

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.

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.

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.

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 Störgeräusch liegen auf den gleichen Harmonischen.

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.

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.

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.

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

- a) Aufteilen eines einkanaligen Sprachsignals in eine Reihe von Frequenzkanälen durch Überführen vom Zeitbereich in einen diskreten Frequenzbereich,
- b) Herausfiltern der Modulationsfrequenz des EL 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.

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).

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.

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.

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.

Die Realisierung des erfindungsgemäßen Verfahrens kann beispielsweise durch ein Software-Plugin, als fest verdrahtete Lösung oder auch als Anlogschaltung erfolgen.

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.

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

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.

In einer weiteren Ausführung des erfindungsgemäßen Verfahrens erfolgt vor der Rücktransformation in Schritt c) eine Gleichrichtung der negativen Signalkomponenten.

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.

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.

Zwar bezieht sich die Erfindung auf die Verbesserung der Sprachqualität eines EL-Sprechers bei Verwendung von elektronischen Mitteln – anstatt eines Zuhörers würden die Signale also beispielsweise mit einem Mikrofon aufgenommen werden. Zur Illustration der Aus-

gangslage wurde allerdings aus Gründen der Verständlichkeit dieses allgemeine Modell gewählt.

Fig. 2 zeigt eine vereinfachte Modelldarstellung der Situation, auf die die erfindungsgemäßen 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.

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

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,

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

Es können nun die Signalteile vom zeitinvarianten und vom zeitvarianten Signalpfad 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 $S_1(w, t) \sim x(w)H(w, t)$.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

ANSPRÜCHE

1. Verfahren zur Verbesserung der Sprachqualität eines Elektro-Larynx (EL) Sprechers, dessen Sprachsignal über geeignete Mittel digitalisiert wird, **gekennzeichnet** durch die folgenden Schritte:
 - a) Aufteilen eines einkanaligen Sprachsignals in eine Reihe von Frequenzkanälen durch Überführen vom Zeitbereich in einen diskreten Frequenzbereich,
 - b) Herausfiltern der Modulationsfrequenz des EL 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.
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.

1/2

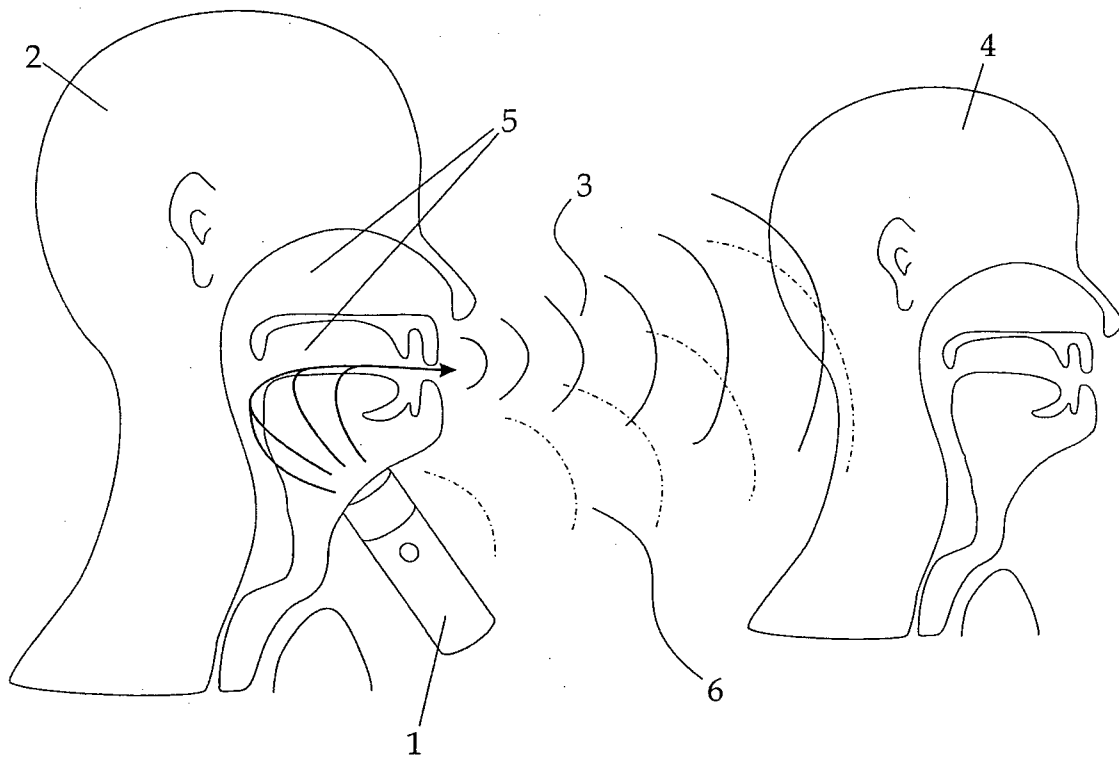


Fig. 1

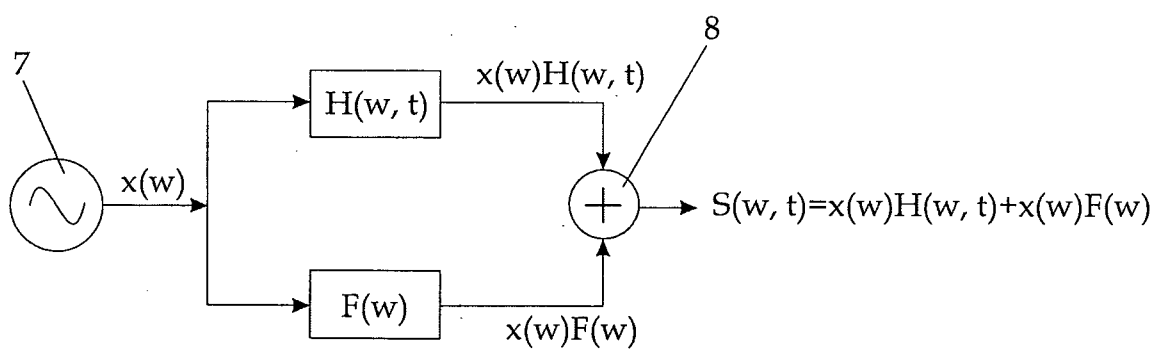


Fig. 2

2/2

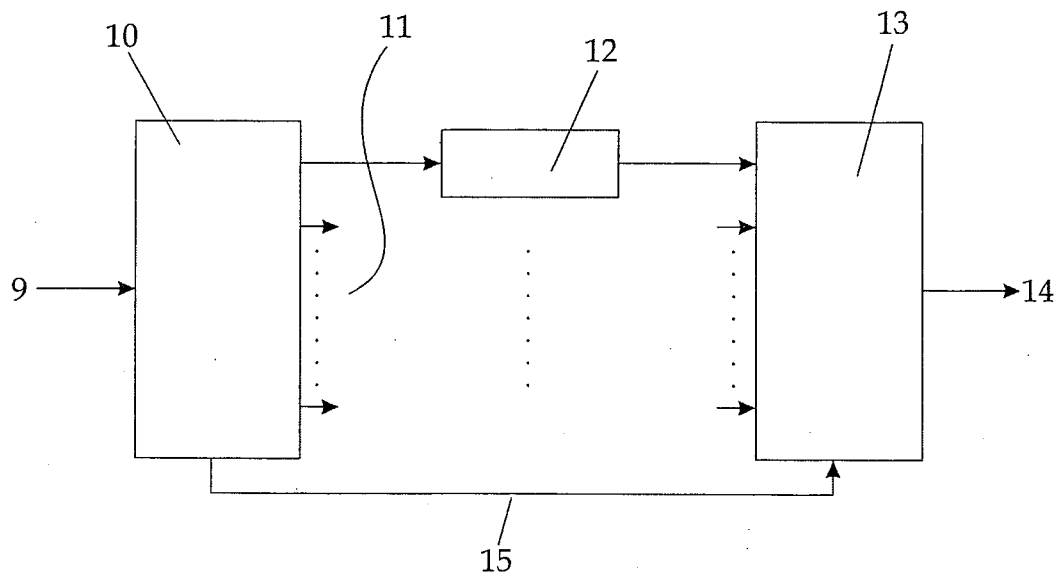


Fig. 3

INTERNATIONAL SEARCH REPORT

International application No

PCT/AT2010/000032

A. CLASSIFICATION OF SUBJECT MATTER

INV. G10L21/02 A61F2/20

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L A61F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	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, vol. 2, 2 December 1997 (1997-12-02), pages 491-494, XP010264286 ISBN: 978-0-7803-4365-8 the whole document ----- -/--	1-5

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *G* document member of the same patent family

Date of the actual completion of the international search

6 July 2010

Date of mailing of the international search report

19/07/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Burchett, Stefanie

INTERNATIONAL SEARCH REPORT

International application No

PCT/AT2010/000032

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HERMANSKY H ET AL: "RASTA processing of speech" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING USA, vol. 2, no. 4, October 1994 (1994-10), pages 578-589, XP002590490 ISSN: 1063-6676 DOI: 10.1109/89.326616 page 579, left-hand column, paragraph 3 - right-hand column, paragraph 5; figure 1	1-5
A	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, vol. 41, no. 6, 1 November 2003 (2003-11-01), pages 670-678, XP001201568 ISSN: 0140-0118 the whole document	1-5
A	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 vol. 23, no. 4, 31 December 2002 (2002-12-31), pages 229-232, XP002590491 Retrieved from the Internet: URL:http://splab.net/papers/2002/2002_02.p df [retrieved on 2010-07-05] the whole document	1-5
A	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, vol. 95, no. 5, PART 01, 1 May 1994 (1994-05-01), pages 2670-2680, XP000447919 ISSN: 0001-4966 the whole document	1-5

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/AT2010/000032

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

INV. G10L21/02 A61F2/20

ADD.

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

G10L A61F

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	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 das ganze Dokument ----- -/--	1-5

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen ☐ Siehe Anhang Patentfamilie

- * Besondere Kategorien von angegebenen Veröffentlichungen :
 - *A* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
 - *E* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
 - *L* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
 - *O* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
 - *P* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
 - *T* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
 - *X* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
 - *Y* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
 - *&* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

6. Juli 2010

Absendedatum des internationalen Recherchenberichts

19/07/2010

Name und Postanschrift der Internationalen Recherchenbehörde

Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Burchett, Stefanie

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	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 Seite 579, linke Spalte, Absatz 3 - rechte Spalte, Absatz 5; Abbildung 1 -----	1-5
A	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 das ganze Dokument -----	1-5
A	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 [gefunden am 2010-07-05] das ganze Dokument -----	1-5
A	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 das ganze Dokument -----	1-5