

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-42740

(P2008-42740A)

(43) 公開日 平成20年2月21日(2008.2.21)

(51) Int.Cl.	F I	テーマコード (参考)
H04R 1/00 (2006.01)	H04R 1/00 327Z	5D015
H04R 3/00 (2006.01)	H04R 3/00 320	5D020
G10L 15/28 (2006.01)	G10L 15/28 400	

審査請求 未請求 請求項の数 4 O L (全 11 頁)

(21) 出願番号	特願2006-217028 (P2006-217028)	(71) 出願人	504143441
(22) 出願日	平成18年8月9日 (2006.8.9)		国立大学法人 奈良先端科学技術大学院大学
(特許庁注：以下のものは登録商標)			奈良県生駒市高山町8916-5
1. Bluetooth		(74) 代理人	100084135
			弁理士 本庄 武男
		(72) 発明者	中島 淑貴
			奈良県生駒市高山町8916-5 国立大
			学法人奈良先端科学技術大学院大学内
		Fターム(参考)	5D015 DD02
			5D020 BB07

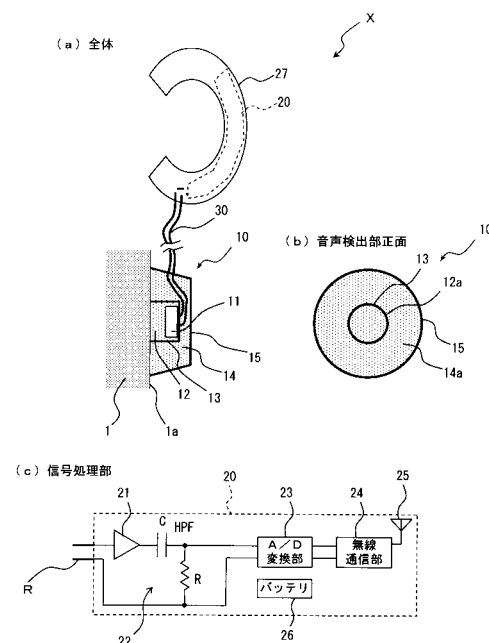
(54) 【発明の名称】 非可聴つぶやき音声採取用マイクロホン

(57) 【要約】

【課題】 ごく簡易な構成により、非可聴つぶやき音声を採取し、これを受話者や音声認識手段が認識しやすい音声信号に変換して出力することができる非可聴つぶやき音声採取用マイクロホン（NAMマイクロホン）を提供すること。

【解決手段】 皮膚表面1aに密着されることにより、人体内を伝播する非可聴つぶやき音声（NAM）を伝播させる軟性部材であるNAM伝播部12、NAM伝播部12を伝播するNAMを電気信号に変換するマイクロホン11、マイクロホン11により得られる音声の信号に対して約1000Hzのカットオフ周波数のハイパスフィルタ処理を施すハイパスフィルタ22とを備えたNAMマイクロホンX。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

人体の声道において生じる声帯の振動を伴わない呼吸音による音声である非可聴つぶやき音声を採取する非可聴つぶやき音声採取用マイクロホンであって、

人体の皮膚表面に密着されることにより前記非可聴つぶやき音声を伝播させる軟性部材と、

前記軟性部材を伝播する前記非可聴つぶやき音声を電気信号に変換するマイクロホンと、

前記マイクロホンにより得られる前記非可聴つぶやき音声の信号に対してハイパスフィルタ処理を施すハイパスフィルタと、

を具備してなることを特徴とする非可聴つぶやき音声採取用マイクロホン。

【請求項 2】

前記ハイパスフィルタのカットオフ周波数が略 800 Hz 乃至略 1400 Hz である請求項 1 に記載の非可聴つぶやき音声採取用マイクロホン。

【請求項 3】

前記ハイパスフィルタのスロープ特性が略 -16 dB/oct 乃至略 -14 dB/oct である請求項 2 に記載の非可聴つぶやき音声採取用マイクロホン。

【請求項 4】

前記マイクロホンと前記ハイパスフィルタとの間で前記非可聴つぶやき音声の信号を増幅するアンプを具備してなる請求項 1～3 のいずれかに記載の非可聴つぶやき音声採取用マイクロホン。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、非可聴つぶやき音声を採取する非可聴つぶやき音声採取用マイクロホンに関するものである。

【背景技術】

【0002】

昨今、携帯電話機及びその通信網の普及により、いつでもどこでも他の人と音声（会話）によるコミュニケーションをとることが可能となっている。さらに、音声認識手段を備えたパーソナルコンピュータやカーナビゲーション装置等の各種の装置が、音声の指令に応じて処理を実行することも可能となっている。

その一方で、電車内や図書館内など、周囲の人への迷惑防止のために発声が制限される状況や、会話の内容が機密事項等であるために発声が制限される状況も多い。そのように発声が制限される状況においても、周囲に発声内容が漏れることなく音声通話や装置に対する音声での指令を行うことができれば、音声によるコミュニケーションのさらなるオンデマンド化や、音声による機器の遠隔制御等が促進され、各種業務の効率化にもつながる。

また、咽頭部（声帯など）に障害があるため通常音声を発声できない障害者であっても、非可聴つぶやき音声であれば発声できる場合が多い。このため、非可聴つぶやき音声による通話や機器に対する指令が可能になれば、そのような咽頭部の障害者の利便性が格段に向上する。

これに対し、特許文献 1 には、非可聴つぶやき音声（NAM：Non-Audible Murmur）を採取することによって音声入力するコミュニケーションインタフェースシステムが提案されている。非可聴つぶやき音声（NAM）は、声帯の規則振動を伴わない音声（無声音）であって、外部からは非可聴な体内軟部組織を伝播する振動音（呼吸音）である。即ち、人体の声道において生じる声帯の振動を伴わない呼吸音による音声である。例えば、防音室環境において、1～2 m 程度離れた周囲の人に聞こえない程度の非可聴音声（呼吸音）を「非可聴つぶやき音声」と定義し、声道（特に、口腔）を絞って声道を通過する空気の流速を上げることにより、1～2 m 程度離れた周囲の人に聞こえる程度に無声音を発声

10

20

30

40

50

する可聴音声を「可聴ささやき音声」と定義する。

このような非可聴つぶやき音声の信号は、音響空間の振動を検知する通常のマイクロホンでは採取できないため、通常、体内の肉伝導音を採取する肉伝導マイクロホンにより採取される。この肉伝導マイクロホンは、従来、主として非可聴ささやき音声（NAM）の採取に用いられるため、NAMマイクロホンとも呼ばれ、その詳細は、特許文献1等に表示されている。

このNAMマイクロホン（肉伝導マイクロホン）は、人体の皮膚表面に密着されることにより非可聴つぶやき音声を伝播させるシリコン等からなる軟性部材と、その軟性部材を伝播する非可聴つぶやき音声を電気信号に変換するマイクロホンとを備えている。そして、NAMマイクロホンは、耳介の下方部における頭蓋骨の乳様突起直下の、胸鎖乳頭筋上の皮膚表面に前記軟性部材が密着するよう装着され、声道で発生して体内の軟組成（骨以外の筋肉や脂肪など）を伝わる肉伝導音（非可聴ささやき音声）を採取する。

ここで、NAMマイクロホンは、主として人体の軟組成を経由して伝播してくる音のみを採取するため、周囲の騒音（ノイズ音）が大きい場合であっても、人体がノイズ除去フィルタの機能を果たし、SN比の高い音響信号を採取することができる。即ち、NAMマイクロホンは、空中を伝播する音響に対する耐ノイズ性が高い。

【0003】

ところで、可聴ささやき音声は、十分な音量に増幅さえすれば、特段の訓練を受けていない一般的な人によってもその発話内容を高い認識率で聴き取ることができる。

一報、非可聴つぶやき音声は、その音声を単に増幅しても、受話者が発話内容を聴き取りにくい（認識率が低い）という問題点がある。このことは、音声認識手段においても同様である。

これに対し、例えば非特許文献1には、統計的スペクトル変換法によるモデルの一例である混合正規分布モデルに基づいて、NAMマイクロホン（肉伝導マイクロホン）により得られる非可聴つぶやき音声の信号を、通常発声した音声（有声音）の信号に変換する技術が示されている。

また、特許文献2には、2つのNAMマイクロホン（肉伝導マイクロホン）により得られる非可聴つぶやき音声の信号のパワーの比較により、通常発声音（有声音）のピッチ周波数を推定し、その推定結果に基づいて、非可聴つぶやき音声の信号を通常発声した音声（有声音）の信号に変換する技術が示されている。

これら非特許文献1や特許文献1に示される技術を用いることにより、体内伝導マイクロホンを通じて得られた非可聴つぶやき音声の信号を、受話者が比較的聴き取りやすい通常音声（有声音）の信号に変換できる。

【特許文献1】WO2004/021738号パンフレット

【特許文献2】特開2006-086877号公報

【非特許文献1】戸田智基 他、「混合正規分布モデルに基づく非可聴つぶやき声（NAM）から通常音声への変換」電子情報通信学会 信学技報、SP2004-107、pp.67-72、2004年12月

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、音声変換モデルに基づいて非可聴つぶやき音声を通常の可聴音声に変換するという従来技術（特許文献1や特許文献2に示される技術）は、その変換に要する時間だけ音声信号伝送の遅延が生じるという問題点を有していた。さらに、その従来技術を実現する装置は、音声変換モデルに基づく信号変換処理を実行する演算手段（マイクロコンピュータ）等が必要となることから、電力消費量が比較的大きく、コストも高いという問題点を有していた。ここで、音声通話や機器に対する音声での指令を行う場合、携帯型或いは身体装着型の音声の入出力装置を構成する必要があるが、その場合、電力消費量が大いという問題は、十分な連続使用時間を確保できないという問題点にもつながる。

また、特許文献2にも示されるように、非可聴つぶやき音声は、声帯の規則振動を伴わ

10

20

30

40

50

ない無声音である。そして、特許文献1や特許文献2に示されるように、無声音である非可聴つぶやき音声の信号を通常音声（有声音）の信号へ変換する場合、声道による音響的な特徴量の変換特性（入力信号の特徴量から出力信号の特徴量への変換特性）を表す声道特徴量変換モデルと、音源（声帯）による音響的な特徴量の変換特性を表す声帯特徴量変換モデルとを組み合わせた音声変換モデルが用いられる。このような音声変換モデルを用いた処理は、声の高さの情報に関して「無」から「有」を作り出す（推定する）処理を含むこととなる。このため、非可聴つぶやき音声の信号を通常音声（有声音）の信号へ変換すると、イントネーションが不自然な音声や本来発声していない誤った音声を含む信号が得られてしまい、受話者や音声認識手段の音声認識率が低下するという問題点があった。

従って、本発明は上記事情に鑑みてなされたものであり、その目的とするところは、ごく簡易な構成により、非可聴つぶやき音声を採用し、これを受話者や音声認識手段が認識しやすい音声信号に変換して出力することができる非可聴つぶやき音声採取用マイクロホン（NAMマイクロホン）を提供することにある。

【課題を解決するための手段】

【0005】

上記目的を達成するために本発明は、次の（1）～（3）に示す構成要素を備えることにより、非可聴つぶやき音声（人体の声道において生じる声帯の振動を伴わない呼吸音による音声）を採用する非可聴つぶやき音声採取用マイクロホン（以下、NAMマイクロホンという）である。

（1）人体の皮膚表面に密着されることにより前記非可聴つぶやき音声を伝播させる軟性部材。

（2）前記軟性部材を伝播する前記非可聴つぶやき音声を電気信号に変換するマイクロホン。

（3）前記マイクロホンにより得られる前記非可聴つぶやき音声の信号に対してハイパスフィルタ処理を施すハイパスフィルタ。

ここで、前記ハイパスフィルタのカットオフ周波数は、例えば、800～1400Hz程度であることが望ましい。さらにその場合、前記ハイパスフィルタのスロープ特性は、 $-16 \sim -14 \text{ dB/oct}$ 程度であることが望ましい。

このように、非可聴つぶやき音声の信号を前記ハイパスフィルタに通して得られる信号（前記ハイパスフィルタの出力信号）は、それを単に増幅するだけで、前記可聴ささやき音声の信号とほぼ同等に聴き取りやすい（認識しやすい）信号になることがわかった。しかも、前記ハイパスフィルタは、ごく簡易で電力消費が極めて少ない回路により実現できる（例えば、コンデンサと抵抗素子とによって実現できる）。

なお、一般に、マイクロホンは音声信号を増幅するアンプを備えているが、本発明に係るNAMマイクロホンにおいては、前記マイクロホンと前記ハイパスフィルタとの間で前記非可聴つぶやき音声の信号を増幅するアンプを備えることが望ましい。

これにより、音量レベルが低い前記非可聴つぶやき音声の信号を、前記ハイパスフィルタの入力信号として十分なレベルに増幅することができる。

【発明の効果】

【0006】

本発明に係るNAMマイクロホンよれば、非可聴つぶやき音声を採用することにより、受話者にとって聴き取りやすい音声信号を得ることができる。その聴き取りやすさは、前記可聴ささやき音声と同程度である。

また、本発明に係るNAMマイクロホンにより得られる音声は、従来手法により得られる通常音声（非可聴音声の信号を、声道特徴量変換モデルと音源特徴量変換モデルとを組合せたモデルに基づいて変換した通常音声（有声音））のように、イントネーションが不自然な音声や本来発声していない誤った音声を含むことがなく安定している。

さらに、本発明によれば、音声変換モデルに基づく信号変換処理（比較的負荷の高い処理）が不要であり、音声信号伝送の遅延が生じず、その処理を実行する演算手段（マイクロコンピュータ）等も不要である。このため、本発明に係るNAMマイクロホンは、携帯

10

20

30

40

50

電話機や身体装着型の機器などの小型の機器に組み込まれるような場合でも、機器の重量や体積の増大や、連続使用時間の短縮を招くことがほとんどない。

【発明を実施するための最良の形態】

【0007】

以下添付図面を参照しながら、本発明の実施の形態について説明し、本発明の理解に供する。尚、以下の実施の形態は、本発明を具体化した一例であって、本発明の技術的範囲を限定する性格のものではない。

ここに、図1は本発明の実施形態に係るNAMマイクロホンXの概略構成図（一部ブロック図）、図2はNAMマイクロホンXが人体に装着された状態を表す模式図、図3は非可聴つぶやき音声の声紋と可聴ささやき音声の声紋とを表す図、図4はNAMマイクロホンXの第1評価実験（自然性評価実験）の結果を表すグラフ、図5はNAMマイクロホンXの第2評価実験（単語認識精度評価実験）の結果を表すグラフである。

【0008】

まず、図1を参照しつつ、本発明の実施形態に係るNAMマイクロホンXの構成について説明する。

ここで、図1（a）は、NAMマイクロホンXの全体構成図、図1（b）は、NAMマイクロホンXの一部を構成する音声検出部10の正面図（皮膚への接触面側から見た図）、図1（c）は、NAMマイクロホンXの一部を構成する信号処理部20の構成を表すブロック図である。なお、図1（a）において、音声検出部10については断面図を示している。

NAMマイクロホンXは、人体の声道において生じる声帯の振動を伴わない呼吸音による音声である非可聴つぶやき音声（NAM）を採取する非可聴つぶやき音声採取用マイクロホンである。

図1（a）に示すように、NAMマイクロホンXは、音声検出部10と、信号処理部20と、それらを接続する信号線30とを備え、信号処理部20は、耳装着用筐体27に收容されている。

音声検出部10は、人体の皮膚表面に装着されることにより、人体の声道によって発生し、体内軟部組織（主として肉の部分）を伝播する非可聴つぶやき音声（NAM）の振動を電気信号に変換するものである。

また、信号処理部20は、音声検出部10による検出信号（音声信号）に対する各種の処理（信号増幅処理や外部機器への伝送処理等）を行うものである。

【0009】

以下、図1（a）及び図1（b）を参照しつつ、音声検出部10について説明する。

この音声検出部10は、マイクロホン11と、NAM伝播部12と、内側カバー部材13と、粘着遮音部14と、外側カバー部材15とを備えて構成されている。

NAM伝播部12は、その一面12aが人体1の皮膚表面1aに密着されることにより、人体内を伝播する非可聴つぶやき音声を伝播させる軟性部材である。このNAM伝播部12は、その音響インピーダンスの特性が、人体における肉部の音響インピーダンスの特性に近い材料、例えば、ウレタンエラストマーやシリコン等により構成されている。これにより、非可聴つぶやき音声を、人体（皮膚）からNAM伝播部12へ効率的に伝播させることができる。

マイクロホン11は、NAM伝播部12（軟性部材）を伝播する非可聴つぶやき音声（振動）を電気信号に変換するものである。このマイクロホン11は、音声の振動を感知する感音部（図1（a）における左側面）の全体がNAM伝播部12に対して直接接触している。これにより、マイクロホン11は、NAM伝播部12の振動（音声）を高感度で検出する。このマイクロホン11の出力信号（非可聴つぶやき音声の信号）は、信号線30を通じて信号処理部20へ伝送される。

内側カバー部材13は、NAM伝播部12の皮膚表面1aとの接触面（以下、皮膚接触面12aという）以外の部分全体を覆うものである。即ち、内側カバー部材13は、一の面が開口状態となった容器状の部材であり、その内側に、マイクロホン11が收容される

とともに、NAM伝播部12（軟性部材）がほぼ隙間なく充填された状態（NAM伝播部12の中にマイクロホン11が埋め込まれた状態）となっている。

粘着遮音部14は、粘着性のある軟性部材であるウレタンエラストマーからなり、内側カバー部材13の外側全体を覆うように形成されている。さらに、粘着遮音部14は、人体1の皮膚表面1aに接触して粘着する部分（以下、皮膚接着部14aという）を有している。この皮膚接着部14aは、NAM伝播部12における皮膚1aとの接触面（以下、皮膚接触面12aという）の周り全体に渡って形成されている。

また、外側カバー部材15は、粘着遮音部14の前記皮膚接着部14a以外の外側全体を覆うものであり、音声検出部10の外装を形成するものである。即ち、外側カバー部材15は、一の面が開口状態となった容器状の部材であり、その内側に、NAM伝播部12及びマイクロホン11を内包する内側カバー部材13が収容されるとともに、粘着遮音部14（ウレタンエラストマー）がほぼ隙間なく充填された状態となっている。

【0010】

ここで、粘着遮音部14は、その皮膚接着部14aが皮膚表面1aに対して粘着することにより、NAM伝播部12の皮膚接触面12aを皮膚表面1aに密着させるとともに、当該音声検出部10を人体に対して接着した状態（装着状態）に保持するものである。なお、NAM伝播部12をウレタンエラストマー（粘着性の軟性部材）によって構成することにより、NAM伝播部12の皮膚接触面12aと皮膚接着部14aとの両方が皮膚表面1aに対して粘着し、当該音声検出部10がより強固に人体に対して保持されるので好適である。

また、粘着遮音部14は、内側カバー部材13の外側全体を覆うとともに、NAM伝播部12の皮膚接触面12aの周り全体で皮膚接着部14aが皮膚表面1aに接着することにより、空气中を伝播する外乱音響がマイクロホン11に浸入することを防ぐ遮音材としても機能する。

【0011】

次に、図1（c）を参照しつつ、信号処理部20の構成について説明する。

信号処理部20は、アンプ21、ハイパスフィルタ22、A/D変換部23、無線通信部24、アンテナ25及びバッテリー26を備えて構成されている。

アンプ21は、音声検出部10から伝送されてくる非可聴つぶやき音声の信号（マイクロホン11により得られる音声の信号）を増幅するものである。

ハイパスフィルタ22は、音声検出部10から伝送されてくる非可聴つぶやき音声の信号（マイクロホン11により得られる音声の信号）に対してハイパスフィルタ処理を施し、処理後の信号を後段の回路へ出力するものである。

ここで、アンプ21は、音声検出部10におけるマイクロホン11とハイパスフィルタ22との間に配置され、ハイパスフィルタ22に入力される前の非可聴つぶやき音声の信号を増幅する。

これにより、音量レベルが低い非可聴つぶやき音声の信号を、ハイパスフィルタ22の入力信号として十分なレベルに増幅することができる。

図1（c）に示すハイパスフィルタ22は、コンデンサと抵抗素子とにより構成されるごく簡易なC-R回路（微分回路）である。例えば、そのコンデンサ容量を0.1 μ F程度、抵抗値を1.6 k Ω 程度とすることが考えられる（カットオフ周波数 1 kHz）。

【0012】

A/D変換部23は、アンプ21により増幅され、ハイパスフィルタ22によりフィルタ処理が施された非可聴音声の信号（アナログ信号）を、所定のサンプリング周波数でデジタル信号に変換するものである。例えば、A/D変換部23は、8 kHz程度のサンプリング周波数でA/D変換を行う。

無線通信部24は、A/D変換部23によってデジタル化された非可聴音声信号を、通信機能を備えた外部装置に対してアンテナ25を通じて無線送信するものである。例えば、周知のBluetoothの通信規格に従って、デジタル音声信号を外部装置に送信する。

バッテリー 26 は、信号処理部 20 を構成する各機器（アンプ 21、A/D 変換部 23、無線通信部 24）に対して電力を供給するものである。また、バッテリー 26 は、音声検出部 10 のマイクロホン 11 が駆動電力を必要とするもの（例えば、バイアス型のコンデンサマイクロホン等）である場合、そのマイクロホン 11 に対しても電力供給を行う。もちろん、マイクロホン 11 が、駆動電力を必要としないもの（例えば、バックエレクトレット型のコンデンサマイクロホン等）である場合には、音声検出部 10 側への電力供給は要しない。

【0013】

図 2 は NAM マイクロホン X が人体に装着された状態を表す模式図である。

図 2 に示すように、音声検出部 10 は、耳介の下方部における頭蓋骨の乳様突起直下の、胸鎖乳頭筋上の皮膚表面に、NAM 伝播部 12 の皮膚接触面 12a が密着するようにして人体に装着（粘着遮音部 14 により粘着）される。これにより、声道で発生した非可聴つぶやき音声（NAM）が、骨等が障害物となることなく体内の肉部から NAM 伝播部 12 へ効率的に伝播する。

また、信号処理部 20 は、その筐体である耳装着用筐体 27 が耳に係合することにより人体に装着される。

これにより、ハンズフリー状態で、音声検出部 10 で検出され、ハイパフフィルタ処理が施された非可聴つぶやき音声の信号が、信号処理部 20 から外部装置に対して無線伝送される。従って、信号処理部 20 と通信可能な外部装置が、例えば、非可聴つぶやき音声の信号に基づく音声認識機能と、認識した音声に応じて自装置の動作を制御する自動制御機能とを備えれば、NAM マイクロホン X の装着者は、ハンズフリーの状態で、かつ、周囲に音が漏れない非可聴つぶやき音声の発声により、外部装置を遠隔制御することができる。

また、信号処理部 20 に、外部装置から非可聴つぶやき音声の信号を受信して音声として出力する機器を付加すれば、非可聴つぶやき音声による通話機となる。この場合、例えば、無線通信部 24 に、外部装置から非可聴つぶやき音声の信号を受信する機能を設ける。さらに、信号処理部 20 に、無線通信部 24 によって受信した音声信号（デジタル信号）をアナログ信号に変換する D/A 変換部と、D/A 変換後の音声信号（アナログ信号）を増幅するアンプと、増幅後の音声信号を出力するスピーカ（イヤホン）とを付加する。これにより、NAM マイクロホン X は、非可聴つぶやき音声による通話機となる。

【0014】

ところで、肉伝導音として採取された非可聴つぶやき音声（NAM）と、空気伝導音として採取された可聴ささやき音声とは、いずれも無声音（声帯の振動を伴わない音声）である。しかしながら、前述したように、非可聴つぶやき音声は、単に増幅しただけではその発話内容を認識し難く（聴き取りにくい）、可聴つぶやき音声は比較的その発話内容を認識しやすい。

図 3 は、従来の NAM マイクロホン（フィルタ処理なし）により採取された非可聴つぶやき音声（NAM）の声紋（a）と、一般的なマイクロホンにより空気伝導音として採取された可聴ささやき音声の声紋（b）とを表す図である。

図 3（a）と図 3（b）とを比較すると、発話内容を比較的認識しやすいささやき音声（b）に比べ、非可聴つぶやき音声（a）は、750 Hz 以下の周波数成分の信号強度が強い傾向がある。これは、非可聴ささやき音声の信号において、750 Hz を超える信号成分が、主として発話内容の識別に寄与する信号成分（S）であり、750 Hz 以下の信号成分が、主として発話内容の識別に寄与しないノイズ成分（N）であることが予想される。

【0015】

以上のことから、肉伝導音として採取された非可聴つぶやき音声の信号から、750 Hz 以下の周波数成分をハイパスフィルタ 22（即ち、ローカットフィルタ）により除去すれば、SN 比が向上することが予想される。

実際に、NAM マイクロホン X により、非可聴つぶやき音声を採取すると、それを単に

10

20

30

40

50

増幅するだけで受話者にとって聴き取りやすい（認識しやすい）音声信号が得られることがわかった。その聴き取りやすさは、前記可聴ささやき音声と同程度である。

また、本発明に係るNAMマイクロホンにより得られる音声は、音声変換モデルに基づいて変換した通常音声（有声音）のように、イントネーションが不自然な音声や本来発声していない誤った音声を含むことがなく安定している。

さらに、ハイパスフィルタ22は、携帯電話機や身体装着型の機器などの小型の機器に組み込まれるような場合でも、機器の重量や体積をほとんど増大させることがなく、また、連続使用時間の短縮を招くこともない。

また、NAMマイクロホンXにおいては、音声変換モデルに基づく信号変換処理のように、演算負荷の高い処理が不要であるので、音声信号伝送の遅延が生じない。これにより、非可聴つぶやき音声によるスムーズな対話や機器の遠隔制御を実現することができる。

10

【0016】

次に、NAMマイクロホンXに採用するハイパスフィルタの特性とマイク性能との関係性を評価した実験である第1評価実験及び第2評価実験について説明する。

第1及び第2評価実験は、いずれも22人の被験者（男性10名、女性12名）が、それぞれ音源や信号処理（ハイパスフィルタ処理）の内容が異なる22種類のサンプル音声（第1～第22のサンプル音声）を聴き取り、その聴き取りによって得た感覚（評価結果）に従って、予め定められた評価項目について回答するという方式で行われたものである。

ここで、第1～第10のサンプル音声は、所定の発話者が、所定のサンプルテキスト（文章、単語を含む）を非可聴つぶやき音声（NAM）で発話したときの肉伝導音を、それぞれカットオフ周波数が200、400、600、…2000Hz（200Hzきざみ）であり、スロープ特性が-120dB/octであるハイパスフィルタを備えたNAMマイクロホンXにより収録されたフィルタ処理後の非可聴ささやき音声である。

20

また、第11～第20のサンプル音声は、所定の発話者が、前記サンプルテキストを非可聴つぶやき音声で発話したときの肉伝導音を、それぞれカットオフ周波数が200、400、600、…2000Hz（200Hzきざみ）であり、それぞれスロープ特性が-23.0、-18.6、-16.8、-15.6、-15.0、-14.4、-13.9、-13.6、-13.2及び-13.0dB/octであるハイパスフィルタを備えたNAMマイクロホンXにより収録されたフィルタ処理後の非可聴ささやき音声である。

また、第21のサンプル音声は、通常のマイクロホン（空気伝導音を採取するマイクロホン）により収録された前記可聴ささやき音声である。

30

また、第22のサンプル音声は、ハイパスフィルタ処理を行わない従来のNAMマイクロホンにより収録された非可聴ささやき音声（NAM）である。

また、前記サンプルテキストは、単語数20～30程度の新聞記事等である。

なお、いずれのサンプル音声も、量子化ビット数が16bit、サンプリングレートが8kHzのデジタル音声（PCM音声）として録音し、これを再生した。

【0017】

<第1評価実験（自然性評価実験）>

第1評価実験では、22種類のサンプル音声の中から任意に選択した2種類で1組（ペア）のサンプル音声を11組ずつ各被験者に聴き取らせ、各組について、いずれのサンプル音声の方が会話音声として自然であると感じたかを選択させた。この第1評価実験では、より多くの被験者によって自然であるものとして選択されたサンプル音声は、自然性が高いといえる。

40

図4は、第1評価実験（自然性評価実験）の結果を表すグラフであり、22種類のサンプル信号それぞれについて、被験者により、当該サンプル信号の方が、比較対象となった他のサンプル信号よりも自然であるとして選択された割合（二者択一選択率）を表す。

図4からわかるように、ハイパスフィルタ処理を施さない非可聴つぶやき音声（第21のサンプル音声v21）は、それが自然であると評価された率（選択率）が5割を下回るのに対し、第15のサンプル音声v15（カットオフ周波数1000Hz、スロープ特性-15.0dB/octのハイパスフィルタ処理）は、それが自然であると評価された率

50

が8割を超えている。ここで、可聴ささやき音声（第22のサンプル音声v22）の選択率が9割強であるので、非可聴つぶやき音声に対して単にハイパスフィルタ処理を施すだけで、自然性が可聴ささやき音声に匹敵する程度まで自然性が向上することがわかる。

また、図4からわかるように、スロープ特性によらず、カットオフ周波数が800～1200Hzのハイパスフィルタ処理を施した非可聴つぶやき音声（第4～第6、第14～第16のサンプル音声信号v4～v6、v14～v16）の選択率が6割を超え、自然性の向上が認められる。

また、図4からわかるように、ハイパスフィルタ処理におけるスロープ特性が急峻（-120dB）な信号よりも、スロープ特性が比較的緩やかな信号（特に、第14～第20のサンプル信号）の方が、自然性の向上に適している。

10

【0018】

<第2評価実験（単語認識精度評価実験）>

第2評価実験では、新聞記事（即ち、意味のある文章）の読み上げ音声である22種類のサンプル音声の中から任意に選択したものを被験者に聴き取らせ、単語の認識精度（被験者が認識した単語の正解率）を評価した。

図5は、第2評価実験（単語認識精度評価実験）の結果を表すグラフであり、22種類のサンプル信号（新聞記事の読み上げ音声）それぞれについて、被験者による単語の認識精度（認識した単語の正解率）を表す。

図5からわかるように、ハイパスフィルタ処理を施さない非可聴つぶやき音声（第21のサンプル音声v21）は、単語認識精度が約85%であるのに対し、第14のサンプル音声v14（カットオフ周波数800Hz、スロープ特性-15.6dB/octのハイパスフィルタ処理）は、単語認識精度が約90%である。

20

また、図5からわかるように、カットオフ周波数が200～1400Hzの範囲で、比較的緩やかなスロープ特性（-23.0～-13.9dB/oct）のハイパスフィルタ処理を施した非可聴つぶやき音声（第14～17のサンプル音声信号v14～v17）は、ハイパスフィルタ処理を施さない非可聴つぶやき音声に比べて同等以上の単語認識精度が得られる。

同様に、スロープ特性が急峻（-120dB/oct）である場合は、カットオフ周波数が200～600Hzのハイパスフィルタ処理を施した非可聴つぶやき音声（第1～3のサンプル音声信号v1～v3）は、ハイパスフィルタ処理を施さない非可聴つぶやき音声に比べて同等以上の単語認識精度が得られる。

30

以上より、前記NAMマイクロホンXにおいて、ハイパスフィルタ22のカットオフ周波数を800～1400Hz程度（特に、800～1000Hz程度）とし、さらにスロープ特性を-14～-16dB/oct程度とすることにより、自然で聴き取りやすい（単語を認識しやすい）音声を採用できることがわかる。

【0019】

以上に示した実施形態では、ハイパスフィルタ処理後の非可聴つぶやき音声を無線伝送（送信）するNAMマイクロホンX（無線タイプのマイクロホン）について示したが、これに限るものではない。

例えば、図1に示した前記NAMマイクロホンXから、A/D変換部23及び無線通信部24が除かれ、その代わりに外部装置の音声信号（アナログ）入力端子に接続する出力端子が設けられた有線タイプのNAMマイクロホンも考えられる。

40

また、ハイパスフィルタ22は、750Hz～1400Hz程度のカットオフ周波数を有するものであれば、CR回路以外の周知の回路により構成されたものも考えられる。

また、ハイパスフィルタ22のスロープ特性は、-16dB/oct～-14dB/oct程度に限るものではない。

【産業上の利用可能性】

【0020】

本発明は、非可聴つぶやき音声を採用するためのマイクロホンに利用可能である。

【図面の簡単な説明】

50

【0021】

【図1】本発明の実施形態に係るNAMマイクロホンXの概略構成図（一部ブロック図）

。

【図2】NAMマイクロホンXが人体に装着された状態を表す模式図。

【図3】非可聴つぶやき音声の声紋と可聴ささやき音声の声紋とを表す図。

【図4】NAMマイクロホンXの第1評価実験（自然性評価実験）の結果を表すグラフ。

【図5】NAMマイクロホンXの第2評価実験（単語認識精度評価実験）の結果を表すグラフ。

【符号の説明】

【0022】

10

X…本発明の実施形態に係る非可聴つぶやき音声採取用マイクロホン

10…音声検出部

11…マイクロホン

12…NAM伝播部

13…内側カバー部材

14…粘着遮音部

15…外側カバー部材

20…信号処理部

21…アンプ

22…ハイパスフィルタ

20

23…A/D変換部

24…無線通信部

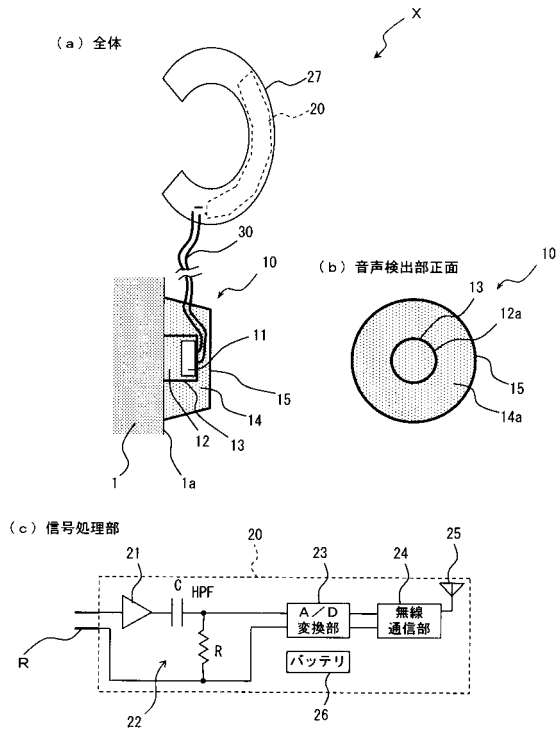
25…アンテナ

26…バッテリー

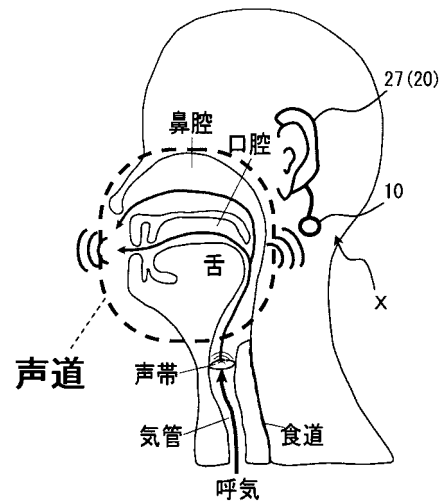
27…耳装着用筐体

30…信号線

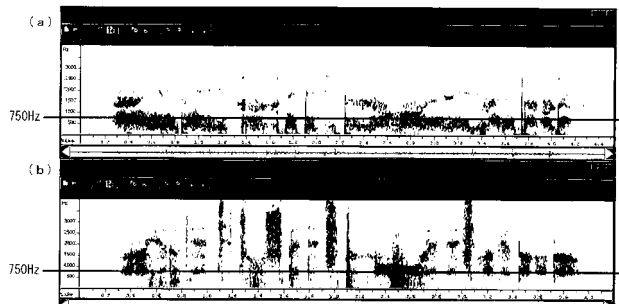
【図 1】



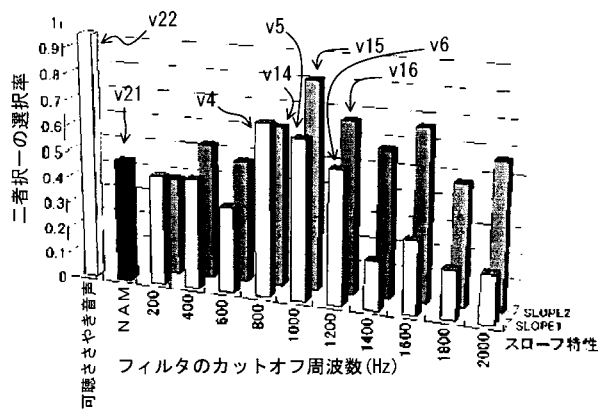
【図 2】



【図 3】



【図 4】



【図 5】

