論　文

非可聴つぶやき認識

中島　淑貴[†]　　柏岡　秀紀[††]　　キャンベル　ニック[††]　　鹿野　清宏[†]

Non-Audible Murmur Recognition

Yoshitaka NAKAJIMA[†], Hideki KASHIOKA[††], Nick CAMPBELL[††], and Kiyohiro SHIKANO[†]

あらまし　「非可聴つぶやき認識」という，新しいスタイルの実用的な入力インタフェースを提案する．これは音声認識の雑音に対する脆弱性，情報の周囲への漏えい性を克服するため，声帯の振動を伴う通常音声の空気伝搬ではなく，「非可聴つぶやき（Non-Audible Murmur: NAM）」，つまり第三者に聴取不能な声帯の振動を伴わない調音呼気音の体内伝導を，体表からサンプリングし，HMM を用いて認識するものである．これを実現するための基礎として，第一に医療用膜型聴診器の原理を応用した体表接着型マイクロホンを開発した．第二として体内を伝導する NAM を採取して認識するために最適な接着位置を発見した．第三として NAM の音響学的性質を検討した．第四として，この部位から採取されたサンプルを用い，HMM 音響モデルに追加学習して NAM 音響モデルを作成した．これらをもとに，日本語ディクテーション基本ソフトウェアを評価に用い，認識エンジン Julius を使用して大語い連続認識実験を行い，NAM 認識の実用可能性を検討した．

キーワード　インタフェース，音声認識，非可聴つぶやき，Non-Audible Murmur（NAM），NAM 認識

1. まえがき

音声認識システムは，人々が思い描いてきた夢であった．ボタンやキーを押すのではなく，人に語りかけるように機械に直接話しかけられたらという思いは，極めて自然であり，SF の世界でも古くより描かれてきたし，実際に約 30 年にわたってその実現が試みられてきた．隠れマルコフモデル（HMM）を用いた大語い連続音声認識（ディクテーション）も可能となり，PC 上のソフトウェアとして安価に販売もされている．認識精度もコマンド認識はいうに及ばず，ディクテーションにおいても十分実用レベルにあるとさえいえる．

しかしこの便利なハンズフリーの入力インタフェースを，日常的に実用している人を周囲に全く見かけないのはなぜであろう．カーナビゲーションシステムでは，その音声認識システムの機能をオフにして使用している人が多い．もしかするとどんなに認識精度が増し，雑音に対して頑健な「使える」音声認識システムが登場したとしても，それがオフィスや日常の生活の場に普及することは困難であるかもしれない．それはその機能面での不足というよりも，音声認識システムが内包する「実用上での本質的な問題」が現在まであまり考えられたことがなかったためである．

音声認識は，その大前提として，外部マイクロホンから，空気中に放散された音を採取して分析する．約 30 年の技術蓄積を経た今でも，その大前提は変わらない．だから本質的に外部雑音，騒音環境に弱い．これは屋外や移動体での使用を前提としたウェアラブル端末での使用を考えた場合，大きな欠点である．

また逆に，オフィスや公共の場での使用を考えた場合，人間の声は大きな騒音源となり，当然これに付随して「入力内容が周囲の人たちに知られてしまう」という欠点がある．現在のようなオフィス環境で音声認識入力を各人が始めたら，入力内容をめいめいが声に出すことになり，大変な騒音環境となる．またそのために誤認識を引き起こす．

加えて，音声認識を使ってみれば，実感として分かるが，機械に向かって声に出して話しかけるのは，第三者にそれを見られると実に「気恥ずかしい」もので

[†] 奈良先端科学技術大学院大学，生駒市
Graduate School of Information Science, Nara Institute of Science and Technology, 8916–5 Takayama, Ikoma-shi, 630–0101 Japan
[††] （株）国際電気通信基礎技術研究所，京都府
Advanced Telecommunications Research Institute International, 2–2 Hikaridai, Seika-cho, Soraku-gun, Kyoto-fu, 619–0288 Japan

ある．特にそれが内容を匿秘したいものであれば，なおさらである．

前述のパラドクスは，これらのことを考えると理解可能と思われる．それに当然ながら，そもそも声帯を振動させる声を出せない障害をもった人々には音声認識は使えない．

個人端末のウェアラブル化，日常生活へのコンピュータやロボットの浸透，世界規模の巨大ネットワークの出現とそのブロードバンド化，無線化．これらのことは音声認識の開発が始まった当時は，現実問題として考えられもしなかった．音声認識はSFで描かれるとおり，ロボットやコンピュータにじかに話しかけることを想定したものだったし，今でも大多数の人にはそう考えられている．

我々は音声認識の普及を妨げている原因の本質が，音声認識のあたり前の大前提として「空気中に放散された通常音声を主に分析対象とし続けてきたこと」にあるのではないかと考えた．上記のような現在や近未来の状況では，むしろ手元の携帯情報端末で，空中放散音声ではなく，発話動態自体を確実に認識し，場合によっては修正してから，ネットワークでテキストなどのパラメータを相手端末や機械に送った方がはるかに現実的である．ただ，せっかく人間が長い歴史の中で育んできた技術習得不能の音声言語文化は，そのまま流用できれば非常に便利で生理的でもある．また長年培ってきた音声認識の素晴らしい技術も生かせれば更に良い．そこで「声を出す」という音声認識や電話の大前提を疑ってみた．

2. 第二の音声言語

人間の音声言語は，声帯を振動させて発生する音源が，調音器官により形成される音響的なフィルタ共振特性によって変化を受けたものを基本としている．無声子音など，声帯の振動を伴わない音素もあるが，ある距離を置いた相手に音声情報を伝達するため，基本的に声帯の振動を伴った有声音を発している．ささやき声は声帯を振動させないが，やはり限定された相手に情報を伝達するため，声門裂を狭めることによって空気の乱流による雑音信号を声帯音源の代わりとしている．どちらも距離の差異こそあれ，他者への音声情報の空気伝達を目的として発声する第一の音声言語である．

ここで我々の日常生活を思い起こすと，人間はもう一つ別の言語発話行動をしていることがある．人に聞かれないように口の中で独り言をつぶやくとき，また祈りや願い事をひそかに口の中で唱えるときである．それはささやき声に似ているが，もっと微弱な，人に聞かれることを前提にしない，むしろ人に聞かれたくない声なき声である．声帯を振動させたり，声門を狭めたりすることもなく，ほぼ呼気に伴って発話器官だけ動かすようなものがこれにあたる．英語のlip thinkと呼ばれるものに近いかも知れない．この発話行動には元来適当な名前が付いていないので，ここではこれを「非可聴つぶやき（Non-Audible Murmur: 以下NAM)」と名づけることにし，厳密な定義は後述する．この発話行動は個人の内部で処理される．また単に思考しているだけではなく，実際に口周囲の運動となって現れる思考の表現の一種でもある．こういう言語活動の存在そのものにさえ気づいていない人もいる．今まではこのNAMがコミュニケーションに使われることはなかった．しかし，その存在に気づきさえすれば，誰もが新たな技術の習得なしに簡単に実行できる，日常的な言語活動である．

このNAMを人対人，人対機械の新たなコミュニケーションインタフェースとして用いることができないであろうか．それを認識して入力インタフェースとして使えないものだろうか．

3. 体表接着聴診器型マイクロホンによる非可聴つぶやき認識

図1は非可聴つぶやき認識の一種として考えられる，NAM認識入力インタフェースの概念図である．皮膚表面に密着して採音されたNAMは，有線，若しくは無線にてマイクアンプやA–D変換器に送られ，実時

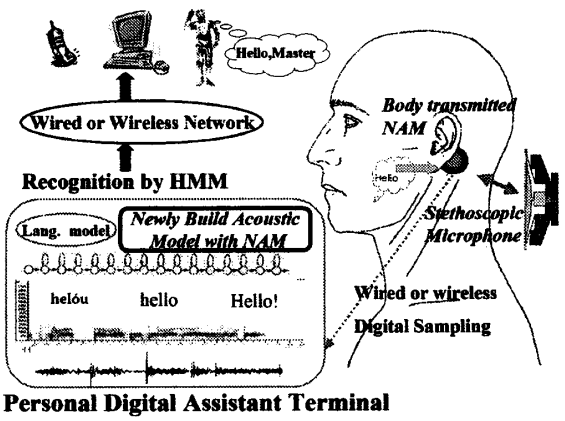


図1 非可聴つぶやき認識の概念図
Fig. 1 A concept chart for NAM recognition.

間音声認識の分析手法として現在最も有力な隠れマルコフモデル（以下 HMM）による分析を行う．この場合認識エンジンは「言語モデル」と「音響モデル」の両方を使用するのが現在の音声認識の主流であるが，この音響モデルについて「通常音声音響モデル」の代わりに新たに「NAM 音響モデル」を作成して置換える．こうすればこの音響モデル置換えの操作以外は，音声認識の従来の技術蓄積をほぼそのまま利用可能である．

大切なことは，人間の音声言語を空中放散させて人や機械に伝達するのではなく，体表や手元で確実に認識分析して，場合によっては視認し，訂正してから，テキストなどのパラメータを相手に電送するという考え方である．現在や近未来に想定される技術やインフラがあれば，それが可能である．

3.1 NAM の概念と定義

非可聴つぶやき（Non-Audible Murmur）という言葉は，「周囲の人に内容を聴取することが困難な，口の中で自己処理的に行う発話行動」を指す造語である．

音響学的には「声帯振動を伴わない無声呼気音が，発話器官の運動による音響的フィルタ特性変化により調音されて，人体頭部の主に軟部組織を伝導したもの」と定義する．

音源として「声門の狭めに伴う乱流（雑音信号）」を用いる無声音の「ささやき声」と NAM とに正確な物理的境界線を引くことは難しい．発話時の口や舌の動きをしながら息を静かに吐き出しただけのものから，「ささやき声」に近いような，声門付近の乱流雑音を音源としているものまで，NAM にもかなりのバリエーションがある．実際 NAM を「聞き取れないほど微弱なささやき声」と考えると，人に説明するときに理解が容易であるし，増幅して聞いた印象も似ている．そういった音の発生源として物理学的，解剖・生理学的な面での境界はあいまいなものの，明確な差異は，「発話意図」と「伝導媒体」である．ささやき声は，明らかに公にはしたくないが，距離的に近い，ある限定した聴者にだけ情報を伝える目的で発せられる音声である．NAM は人に聞こえないように，または聴者を想定せず，自分の中で処理される発話行動である．NAM は体表に付けた特殊なマイクで検出できる程度のパワーの低い「ささやき声」であるともいえる．特にまた「ささやき声」をはじめとする音声は，もちろん空気伝導であり，現在までの研究でも，常に外部マイクによる採音収録を想定している．この点で人間の軟部組織，つまり肉の振動が伝導したものであると定義した NAM とは本質的に異なる．

また他にも，口だけを動かす，いわゆる「口パク」という類発話行動もあるが，これは呼気の全く伴わない NAM であるともいえる．しかし全く呼気を伴わない発話行動というのは，実際にやってみれば理解できるが，実用上かえって不便で難しく，無論長い文章は発話できないし，発話，無発話のオン・オフもあいまいとなる．

周囲の人に聞かれることや音声コミュニケーションを目的とせず，自己内部の言葉として自然に無声音で発話したものが NAM といえるかも知れない．また調音器官の運動を行いながら息を静かに吐き出したものともいえる．図 2 に NAM の概念の図示を試みた．

3.2 体表接着聴診器型マイクロホンの開発

医療用膜型聴診器を顎の下に当ててみたときに，周囲や自分にも聞こえないはずの小さなささやきが聞き取れることの発見があったため，NAM マイクロホンは当初，医療用聴診器を切断して，中にコンデンサマイクロホンを埋め込むことから始まった．

図 3 は音響モデルを実際に作成した NAM マイクロホンのモデルである．市販の粘着面のある吸盤用ポリエステル固定板（40 mm 径）と，エラストマー樹脂吸盤とを組み合わせた．固定板が振動板も兼ねており，

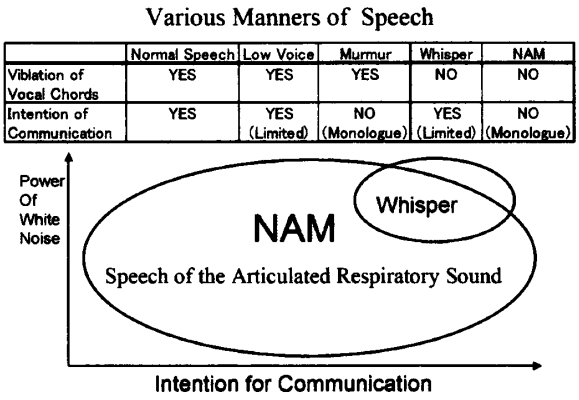


図 2 NAM の概念
Fig. 2 Concept of NAM.

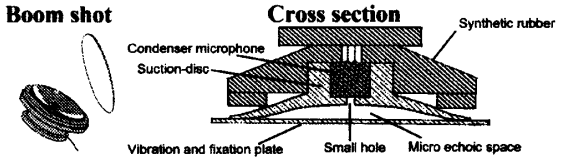
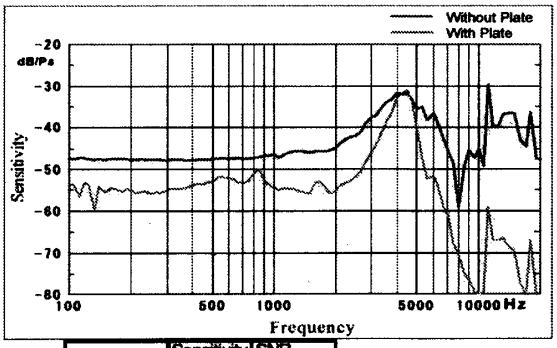


図 3 実際に音響モデルを作成した NAM マイクロホンの構造
Fig. 3 Structure of NAM microphone.



	Sensitivity	SNR
Without Plate	-46.43	56.92
With Plate	-54.34	49.01

図 4　NAM マイクロホンの周波数特性（非装着時）
Fig. 4　Frequency responses of NAM microphone.

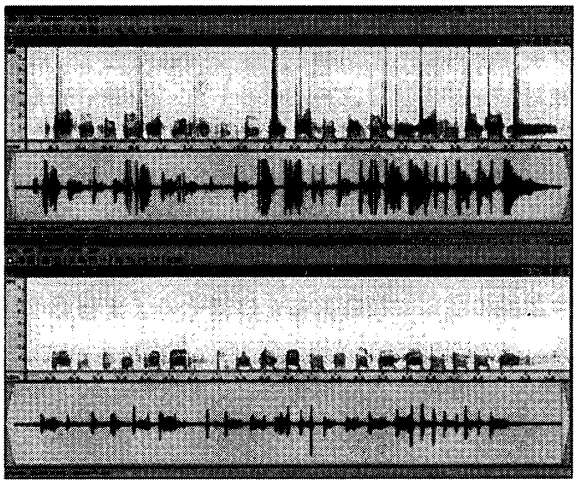
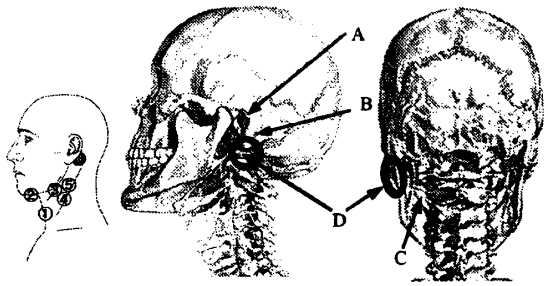


図 5　一般的な NAM の音声波形とスペクトル
Fig. 5　Waveform and spectrum of general NAM.



A: Earhole　B: Mastoid Process　C: Opening of the Mouth
D: ***The Best Sensing Position for NAM Recognition***

図 6　NAM マイクロホン接着位置
Fig. 6　Location of the NAM microphone attachment.

しかも吸盤で固定しながら，膜型聴診器の原理である微小密閉反響空間を作り出せるという一石二鳥の効果をねらった．マイク裏面の防音には AV 機器固定用の弾まない合成ゴムを使用した．部品はすべてホームセンターなどで入手可能であり，原価は千円以下と低コストで非常に軽量である．また装着感は，慣れれば音楽用ヘッドホンより気にならない．24 時間装着していてもはずれることはなかった．

このような NAM マイクロホンは容易に大量生産可能であり，現在のエレクトロニクス技術をもってすれば，マイクアンプやワイヤレス通信回路を小さく基板化して，内部に埋め込むことも十分可能である．また携帯電話のように充電器に置いておくだけで，充電されるようなシステムも考えられる．

図 4 に NAM マイクロホンの周波数特性を掲げておく．4〜6 KHz で 15 dB 程度の大きなピークが生じているが，これは適度な柔らかさと弾性をもつ吸盤によって図 3 の様な聴診器構造をとったときに見られる特性である．

図 5 はこの開発した聴診器型 NAM マイクロホンを，左下顎の耳下腺付近に接着して採取した NAM の音声波形とスペクトルであり，発話内容は「かきくけこたちつてとぱぴぷぺぽばびぶべぼ」である．上段ではゲインを上げて，下段ではゲインを下げて収録したサンプルである．

3.3　NAM マイクロホン最適接着位置の発見

上記 NAM マイクロホンとマイクアンプを用い，下顎の耳下腺部付近や側頸部の皮膚からサンプリングして，音響モデル作成の予備実験を行った．しかし母音は比較的良好に認識するが，子音の判別が困難であるという結果しか得られなかった．入力ボリュームを様々に変化させても，子音のパワーが母音に比して強いため，母音の第 1，第 2 フォルマントがある程度明確に描出されるようにして弁別を良くしようとすると，図 5 の上段のように子音の音声信号がオーバフローしてしまい，摩擦音や破擦音もすべて同じ破裂音として聴取された．また図 5 の下段のように，破裂音などの子音を中心に考えてゲインを下げると，母音は小さすぎてふめいりょうとなるジレンマに陥った．これは NAM マイクロホンが人間の肉の振動を直接拾うため，肉同士が接触したり摩擦したりすることの多い子音のパワーが，音響管の共鳴である母音のパワーよりも相対的に強くなるからであると考えた．

そこで耳下腺の下顎角に近い部位に接着していたものを，図 6 の左図の番号のごとく移動して採音してみた．だが子音・母音のパワー比が認識に適した部位を特定できず，数値評価するに足る認識率は得られなかった．頸部は頸動脈の拍動の雑音が混入した．前頸部では母音のフォルマントの差異が小さくなった．しかし，図 6 の二重丸に示したように，耳に近い高い位

置に接着すると，子音・母音パワー比の近い，認識に適すると考えられる音声波形とスペクトルが得られた．

図 6 の解剖図のごとく，頭蓋底の耳孔のすぐ後に，乳様突起と呼ばれる骨の突起が存在する．これは大きな首の筋肉（胸鎖乳突筋）と頭蓋骨とをつなぐ起始部となる部位である．これに振動板の上部が一部かかるような位置にマイクを接着すると，NAM の子音・母音パワー比は認識に至適となる．実際採音されたものを試聴すると，人間にも認識しやすい．またこの位置は太い筋肉の上にマイクが乗り，大血管の拍動などの振動雑音も入らない．加えて耳の後ろのこの部位は，頭髪と髭の境界であって，日本人では特に毛髪の生えない皮膚のむき出しになった部位であり，女性でも髪を上げると，この部位は無毛で，実用面でも接着位置に最適である．また乳用突起という骨の先端に固定板の一部がかかるので，固定は一段としっかりする．しかし振動板の中心部は筋肉上で，解剖学的に見ても，この部位は調音器官である声道を，上は頭蓋底，左右を下顎骨と頸椎に挟まれた骨の間の窓を通して，斜め後ろ側からまさに水平に眺めた形になる．骨などの音響的障害物なしに，筋肉や結合組織など，ほぼ同じ音響インピーダンスの軟部組織のみを通して，直線的に見渡せる構造となっていて，調音器官の共鳴による音響フィルタ特性をとらえるに適している．しかもある種の子音が作り出す肉の摩擦や接触からはある程度の距離がある．

これより上の頭蓋骨にあたる部分に装着すると，音声波形そのものの振幅が小さくなり SN 比が劣化した．

加えて偶然ではあるが，この部位はウェアラブルな眼鏡型出力デバイスが普及したとすれば「眼鏡の柄」の終点にあたる．また最近流行の耳掛け式ヘッドホンの耳介への固定部の終点でもある．

なお今回自作した肉伝導の聴診器型 NAM マイクロホン以外に，NAM がクリアにサンプリング可能ならば，もちろん市販の圧電素子を使用する骨伝導マイクロホンを使用してもよいのであるが，入手できた耳孔式の骨伝導マイクロホンでは，通常音声は比較的クリアに採取できたが，NAM についてはパワーが小さすぎて余りにも SN 比が低く使用できなかった．

図 7 に上記最適位置から NAM マイクロホンにてサンプリングした NAM のスペクトル，基本周波数 F0，音声波形を並べて掲示する．発話内容は「あらゆる現実をすべて自分の方へねじまげたのだ」である．NAM は声帯振動を伴わない無声音であることが，基本周波数 F0 にプロットを認めないことから分かる．他部位から採取した図 5 のごとき一般的な NAM と比較して，より母音と子音のパワー比が近くなっている．

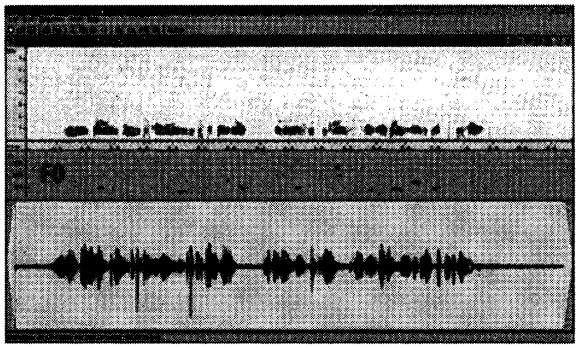


図 7　最適位置からサンプリングした NAM
Fig. 7　NAM from the best sensing position for recognition.

3.4　NAM の音響学的性質

NAM を増幅して耳で聴いた印象は，「やや音のこもったささやき声」であり，図 7 のごとく記録された NAM であれば，内容はほぼ聞き取ることが可能である．そのまま通信に使用したとしても，コミュニケーションの目的は達することが可能であると思われた．

マイクロホン特性とゲインを同条件にして比較するために，音響モデルを作成した NAM マイクロホンとマイクアンプにて，通常音声，ささやき声と NAM をサンプリングした．その際，口唇より 5 cm 離して空中伝導音をとらえたものと，最適接着位置に装着して体内伝導音をとらえたものを，音声波形，スペクトル表示し，並列して図 8 と図 9 に掲げる．左列が体表接着収録，右列が空中収録で，発話内容はすべて「あいうえお，あかさたなはまやらわ」である．明確な境界線を引きにくい常識的な「ささやき声」と「実際音響モデルを作成した NAM」との量的な比較のために，近傍にいる人に内緒事を伝えるときの通常の音量であると思われる一般的な「ささやき声」も収録した．

一般に同音量で発声しても，体内伝導音声の方が，約 5 cm 距離の空気伝導音声よりも波形の振幅は大きくなる．NAM とささやき声に物理的な境界線は引くことが難しいと前述したが，強いていえば，NAM の場合は「非可聴つぶやき」の名前のごとく，伝達の意図をもって発話しないので，図 8 の右 1 段目のように，空中での波形振幅がほとんどなくなってしまう．ささやき声の場合は呼気流量が大きく，声門裂及びその上部構造の狭めが強いため，乱流雑音の音量が大きいためである．

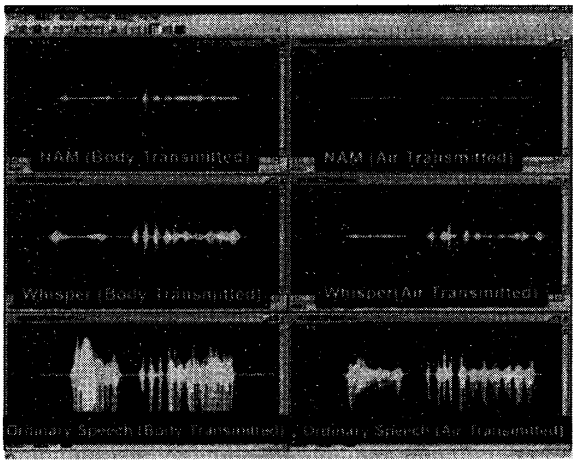


図 8 NAM，ささやき声，通常音声の音声波形
Fig. 8 Waveforms of NAM, whisper and normal speech.

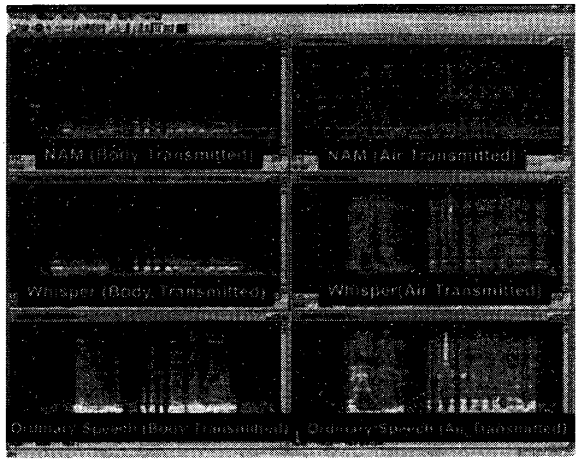


図 9 NAM, ささやき声，通常音声のスペクトル
Fig. 9 Spectra of NAM, whisper and normal speech.

NAM のスペクトル描出の際，約 2 kHz 以下の第 2 フォルマントまでは描出されるが，第 3 フォルマント以上は描出されない．図 4 の NAM マイクロホン特性に見られるごとく，むしろ 2〜4 kHz の感度が良いにもかかわらずである．これは声道から軟部組織を振動が伝導するため，信号は低域フィルタを通過した状態となることと，NAM マイクロホンを音響管の途中に置いているため唇からの放射特性の影響がないためであると考えた．

3.5 NAM 音響モデルの作成

NAM を認識可能な音響モデルを作成するには次の三つの方法がある．

- はじめから NAM サンプルのみを用いて NAM 専用の音響モデルを作る．
- 不特定話者通常音声モデルに多数の NAM サンプルを追加学習する．
- 不特定話者通常音声モデルに少数の NAM サンプルで MLLR [4] などの手法により話者適応する．

上のものほどモデルは NAM に特化されたものとなるが，データ収集や操作が煩雑になる．NAM は増幅すれば「音のこもったささやき声」に近い声としてそのまま聞き取れるほどではあるが，そのスペクトルの様相が余りにも通常音声やささやき声と異なる印象を受けたため，NAM 音響モデル作成としては初めての試みとして，ここでは 2 番目の不特定話者モデルに追加学習を行う方法をとった．

ケンブリッジの HMM ツール集である HTK [6] と，IPA の日本語ディクテーション基本ソフトウェア [4] を用いてこれを行った．

学習サンプル文は特定一個人の NAM にて図 3 の NAM マイクロホンを図 6 の最適位置（左側）に接着し，室内静環境で NAM にて読み上げた．用いた文章は ATR 音素バランス文（A〜J の 503 文＋ Z22 文の計 525 文）を 4 回と JNAS（日本音響学会新聞記事読み上げ音声コーパス）の毎日新聞記事 1255 文を 2 回である．マイクアンプは既述のもの，計算機は linux+ALSA ドライバの環境で，サンプリング周波数は 16 kHz，16 bit にて収録した．

特徴パラメータ抽出は，通常音声と同様の条件で，MFCC（12 次元）＋ Δ MFCC ＋ Δ LogPow（計 25 次元）にて Hcopy [6] により音響分析した．

音素ラベルは時間情報なしのものを用い，HERest [6] にて日本語基本ディクテーションソフトウェア CD-ROM 付属のモノフォン男性不特定話者モデル（状態数 5，混合数 16）を初期モデルとして追加学習を行った．

3.6 認識実験と結果

認識エンジンは上記 CD-ROM 付属の julius を用い，音響モデルを変更する以外の条件は通常音声の認識と同じとし，システムの設定なども特に変更しなかった．言語モデルでは付属の 20 K 辞書を使用した．

評価はいずれも日本語ディクテーション基本ソフトウェア付属の正解文ファイル seikai.ref に記述された毎日新聞記事 24 文を上と同条件で別に録音し，やはり付属の mkhyp.pl，align.pl，score.pl の三つの Perl スクリプトにて認識率を集計した．

結果が表 1 である．なお各環境の内訳は以下のとおり．

A：鉄筋のマンション内の静音環境．

表 1　NAM の大語い連続認識実験
Table 1　Results of the large vocabulary continusous speech recognition of NAM.

Env.	Snt	Corr	Acc	Sub	Del	Ins	Err	S.Err
A	24	93.6	93.3	4.72	1.67	0.28	6.67	50.0
B	24	91.1	90.0	6.67	2.22	1.11	10.0	62.5
C	24	89.7	89.2	9.17	1.11	0.56	10.8	66.7
D	24	90.4	88.4	7.85	1.74	2.03	11.6	60.9

(Env.:録音環境，Snt.:発声文数，Corr.:単語正解率，Acc.:単語認識精度，Sub.:置換誤り率，Del.:脱落誤り率，Ins.:挿入誤り率，Err.:誤り率，S.Err.:文誤り率)

B：ステレオ音響のクラシック音楽を通常楽しむ音量でかけた同室内．
C：NHK のテレビニュースを内容を聞き取るために十分な音量でかけた同室内．
D：診療所の外来で，職務上の音声や人の行き交う音，待合室の静かな会話は聞こえる．仕事中のオフィス内にほぼ相当すると思われる．

まず静音環境では，特定話者モデルながら，モノフォンモデルにもかかわらず，単語認識精度が 90%を超えた。

また日常室内で経験する BGM やテレビの音声などにも頑健であり，B～D に見られるように日常生活空間内や通常のオフィス環境程度の雑音ならば，ほぼそれに劣らず 90%前後の認識精度を示した．

ただし今回の聴診器型 NAM マイクロホンでは，側背部の防音が完全でないのと，コンデンサマイクロホンの入力ゲインを上げているため，採取した雑音環境サンプルに若干の外部雑音が混入しており，これが B～D の認識率をやや低下させたと思われる．そのほかに人間の体自体を伝達する外部雑音もある．

4. むすび

音声認識入力が内包する日常的普及への本質的欠点と，いわゆる「無音声認識」の実用的価値を考察した．

かつてコミュニケーションや入力の方法として使われたことのなかった，調音呼気音の体内伝導を「非可聴つぶやき（Non-Audible Murmur: NAM)」として定義し，これを第二の音声言語として，通常の音声認識と同様に認識したり，加工したりすることにより，人対機械，また人対人の新たなコミュニケーションインタフェースとする可能性を提示した．

聴診器の原理に基づいて，NAM マイクロホンを開発し，最適接着位置を発見し，NAM 音響モデルを作成することによって，NAM による大語彙連続認識の実験を行った。通常音声認識と比較して，その大きな特徴は「人に聞こえないこと」,「体表から直接センシングすること」,「外部雑音に対して頑健であること」などである．

この入力方式が，通常音声で発話できないハンディキャップをもった人々への大きな力となることが期待される．またこの NAM 認識は携帯端末がウェアラブル化されたとき，キーボードやテンキーに代わってその入力の主力となる可能性を秘めていると考える．また第一の音声を人間本来の会話に，NAM という第二の音声を，機械との会話や，人との遠隔コミュニケーションにという使い分けも可能となるかもしれない．

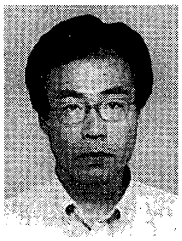
それは音声認識の長く，たゆまない技術蓄積のもとに初めて可能となるものであり，その実用化にも音声認識の研究で培った多くの技術をそのまま生かすことができる．また逆に非可聴つぶやき認識の実用化が，音声認識技術自体の広範な日常的普及の一助となるばかりでなく，音声言語を扱う科学技術に貢献できると考える．

文　献

[1] M. Matsuda, H. Mori, and H. Katsuya, "Formant structure of whispered vowels," J. Acoust. Soc. Jpn. (E), vol.56, pp.447–487, 2000.
[2] L. Rabiner and B.H. Juang, Fundamentals of Speech Recognition, PTR Prentice-Hall, New Jersey, 1993.
[3] H. Suzuki, "Pitfalls when using microphones," J. Acoust. Soc. Jpn. (E), vol.55, pp.377–381, 1999.
[4] T. Kawahara, A. Lee, T. Kobayashi, K. Takeda, N. Minematsu, S. Sagayama, K. Itou, M. Yamamoto, A. Yamada, T. Utsuro, and K. Shikano, "Overview of Japanese dictation toolkit 1999 version," J. Acoust. Soc. Jpn. (E), vol.56, pp.255–259, 2000.
[5] P. Monson and W.R. Emlin, Quantative study of whisper, pp.53–65, Folia Phoniatr. 36, Publisher, 1984.
[6] S. Yong, J. Jansen, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, The HTK Book, Cambridge University Engineering Department, 2000.
[7] 鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄，IT Text 音声認識システム，オーム社，東京，2001.
[8] 中島淑貴，柏岡秀紀，鹿野清宏，キャンベル　ニック，"微弱体内伝導音抽出による無音声認識，" 音響春季講論集，3-Q-12, pp.175–176, March 2003.
[9] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell, "Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin," Proc. ICASSP, pp.708–711, 2003.
[10] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell, "Non-audible murmur recognition," Proc. EUROSPEECH, pp.2601–2604, 2003.

[11] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, and K. Shikano, "Accurate hidden Markov models for non- audible murur (NAM) recognition based on iterative supervised adaptation," Proc. ASRU, pp.73–76, 2003.

（平成 16 年 2 月 16 日受付，4 月 30 日再受付）



中島　淑貴　（学生員）

昭 62 東大・医・医学卒．平元～3 帝京大医学部法医学教室助手．平 3～7 消化管内視鏡を専門に，内科医として民間病院にて臨床に従事．平 7～12 医療法人いわき済生会松村総合病院救急医療センター勤務．平 13 奈良先端科学技術大学院大学情報科学研究科前期博士課程入学．平 15 同卒業．現在，同大学院後期博士課程在学中．また医療法人爽神堂七山病院に内科医として勤務．平 15 日本音響学会ポスター賞．日本医師会認定産業医，日本内科学会，内視鏡学会，日本音響学会各会員．

柏岡　秀紀

1993 大阪大大学院基礎工学研究科博士後期課程了．博士（工学）．同年 ATR 音声翻訳通信研究所入社．1998 同研究所主任研究員（現 ATR 音声言語コミュニケーション研究所）．1999 奈良先端科学技術大学院大学情報学研究科客員助教授．主に自然言語処理，機械翻訳の研究に従事．言語処理学会，情報処理学会，人工知能学会，日本認知科学会各会員．

キャンベル　ニック

英国サセックス大にて Ph.D 取得．リサーチフェローとして IBM 英国科学センターに招聘され，音声合成アルゴリズム開発に従事．AT&T ベル研究所にて客員研究員．エジンバラ大学の言語技術研究センターの Senior Linguist として従事し，1990 年 ATR に移籍．特に大規模音声データベース，波形接続型音声合成，韻律情報モデリングなどに興味をもつ．奈良先端科学技術大学院大学、神戸大学の客員教授を兼任．現在 ATR にて、ネットワーク情報科学研究所のコミュニケーション創発研究室に勤務し，JST/CREST「表現豊かな発話音声のためのコンピュータ処理システムプロジェクト」の主幹研究員．

鹿野　清宏　（正員）

昭 45 名大・工・電気卒．昭 47 同大大学院修士課程了．同年電電公社武蔵野電気通信研究所入所．昭 59～61 カーネギーメロン大客員研究員．昭 61～平 2 ATR 自動翻訳電話研究所音声情報処理研究室長．平 4 NTT ヒューマンインタフェース研究所主席研究員．平 6 より奈良先端科学技術大学院大学情報科学研究科教授．音情報処理学講座を担当．工博．主として音声・音情報処理の研究及び研究指導に従事．昭 50 本会米沢賞，平 3 IEEE SP 1990 Senior Award，平 6 日本音響学会技術開発賞，平 12 情報処理学会山下記念研究賞，平 13 VR 学会論文賞．IEEE，ISCA，情報処理学会，音響学会，VR 学会各会員．