

非可聴つぶやきをインタフェースとするコミュニケーションのための  
ソフトシリコーン型 NAM マイクロホンの開発

中島 淑貴<sup>†</sup> 鹿野 清宏<sup>†</sup>

Development of Soft Silicone Conduction Type NAM Microphones  
for Communication Using Non-audible Murmur as an Interface

Yoshitaka NAKAJIMA<sup>†</sup> and Kiyohiro SHIKANO<sup>†</sup>

あらまし 人体に直接装着して非可聴つぶやき (Non-Audible Murmur : NAM) 音をサンプリングするための肉伝導音センサである次世代 NAM マイクロホンを開発した。これは皮膚に接触する音媒体として、人間の軟部組織に近い音響インピーダンスをもったソフトシリコーンを用いたものである。音声の生成系である人体表面から直接 NAM 音をサンプリングすることにより、より高感度で NAM を聴取可能な音声信号としてとらえることが可能となり、同時に気導外部雑音は人体にフィルタリングされて低減することを、気導音と肉伝導音のステレオ同期同ゲイン収録により示した。ソフトシリコーン型 NAM マイクロホンは従来の聴診器型に比し、NAM においても体内伝導通常音声 (Body Transmitted Ordinary Speech : BTOS) においても帯域が大きく広範化したことによって、HMM による機械認識の精度が向上したばかりではなく、人による聞き取り試験で文章聞き取りの精度が上昇した。しかし肉伝導音声の直接聴取では、無意味単語の聞き取り精度が気導通常音声に比して低い。これはいわゆる「無音声電話」の開発など、NAM を汎用音声入力インタフェースとして用いるに当たった課題である。

キーワード インタフェース, 肉伝導, 非可聴つぶやき, ソフトシリコーン伝導型 NAM マイクロホン

1. ま え が き

携帯電話がこれほどまでに浸透し、ユビキタス社会が声高に叫ばれる今、その入力インタフェースは大きな課題であり、ユビキタスという意味からして、その環境との関係性を考慮せずには語れない。周囲の人々に聞かれることなく、また雑音として迷惑をかけることなく、音声を入力インタフェースとして使えたらと感じたことはないだろうか。

非可聴つぶやき (Non-Audible Murmur : 以下 NAM) は、「周囲の人が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無声音の発話行動」を指す造語である。音響学的には NAM 音を「声帯振動を伴わない無声呼吸音が、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 (肉伝導) したものと定義する [8]。つまり肉伝導・微弱・調音無声音のことである。我々は自作したセンサ (聴診器型 NAM マイクロホン) を耳介後下方部、すなわち乳様突起直下の胸鎖乳突筋上の皮膚に装着することにより、気導音としては周囲に聴取困難なこの NAM を、通常音声サンプリング程度の増幅率で収録することに成功した。このデータを用い、通常音声音響モデルに追加学習や話者適応を行うことによって NAM 音響モデルを作成し [18]、認識エンジン Julius [6] を用いて NAM による HMM 大語彙連続認識が可能であることを示した [8]~[10], [12]。

非可聴つぶやき認識 (NAM 認識) による、いわゆる「無音声認識」の可能性に加えて、NAM 音は増幅のみで他の信号処理をせずに聞いても、聞き慣れればある程度聞き取り可能である。よって声質変換や音源付与などの信号処理技術により、聞き取りやすい音声に変換して通信に使うことも可能であり、いわゆる「無音声電話」等の応用が考えられる。先の「無音声認識」と併せて、人対機械ばかりではなく、人対人の

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科, 生駒市  
Graduate School of Information Science, Nara Institute of  
Science and Technology, 8916-5 Takayama, Ikoma-shi, 630-  
0101 Japan

コミュニケーションの汎用入力インタフェースとして NAM が使用できる。これが実現すれば、周囲の状況に気兼ねする必要のない、しかも外部雑音に対して頑健な、環境に極めて影響されにくい新しい音声入力インタフェースとなる可能性を秘めている。

それを実現するためには、認識のためにも、通信のためにも、未信号処理の NAM 音信号の音質を可能な限り向上させて情報量を増やし、判別性を上げる事が重要課題である。今回我々は、体表密着型肉伝導音センサとして極めて効率の高い、新型の NAM マイクロホンを開発したので報告する。

## 2. 聴診器型 NAM マイクロホン

NAM による大語彙連続認識が可能であることを示した NAM マイクロホンは、図 1 のごとく医療用聴診器の原理を応用して自作したものであった [9]。内蔵されたコンデンサマイクロホン（以下 ECM）と振動板との間の、円錐形の微小密閉反響空間が、軟部組織を伝導する音の感度を上げるのに重要な役割を果たしている。振動板は片面が粘着性で、固定板も兼ねており皮膚に接着する構造となっている。

NAM マイクロホンで収録した通常音声（Body Transmitted Ordinary Speech : 以下 BTOS）と定義する。この聴診器型 NAM マイクロホンでは、NAM 信号についても BTOS 信号についても、医療用聴診器と同様、2kHz 付近に急激なカットオフが見られ、これ以下にしかスペクトルにフォルマントが描出されない。図 3 と図 4 の最上段は聴診器型 NAM マイクロホンによる NAM 音と BTOS 音である。未信号処理の NAM 音を増幅して聞いた印象は「非常に音のこもったささやき声」に近いものであり、初めて聞いて会話内容を聴取することは困難である。特に高域の摩擦性の子音などを含む音声が聞き取りにくい。BTOS 音についても、判別性は NAM よりは良いが、「非常に音のこもった通常音声」であった。

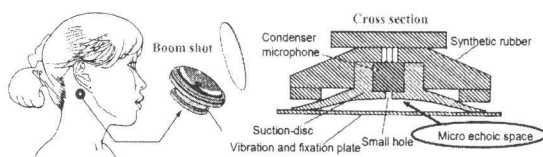


図 1 聴診器型 NAM マイクロホン  
Fig. 1 Stethoscopic NAM microphone.

## 3. ソフトシリコン型 NAM マイクロホン

聴診器型 NAM マイクロホン内部の円錐形状の微小密閉反響空間は、医療用聴診器と同じ原理で低域の感度を上昇させる利点もある反面、この空気層が逆に帯域を制限しているのではないかと考えた。そこで振動板とコンデンサマイクロホン（Electret Condenser Microphone : 以下 ECM）との間に介在する音媒体としての密閉空気層を排し、新たに音媒体として、弾性があり形成が容易な歯科技工用のシリコンを選択した。また ECM は元来、気導音を採録するために設計されているため、ECM 表面に開けた小孔から空気の振動を振動電極板に伝える構造となっている。我々は図 2 の左上のごとく、ECM の表面金属を削り取り、振動電極を完全に露出した形態の ECM を作成した。これを Open Electret Condenser Microphone（以下 OEMC）と名づける。この OEMC を硬質の消しゴム程度の硬さのハードシリコンに完全に包埋し、接着剤を使用せず直接皮膚に圧着したところ、聴診器型に比し感度はやや落ちるものの、2kHz 以上の帯域もやや表現されるようになった。

二つの物質の音響インピーダンスが異なれば、その物質間の接触境界面で音の反射が起こる。医療用超音波イメージング装置を用いて人体の内部構造の観察が可能であるのは、この原理を利用している [14]。我々は超音波プローブと皮膚との間に様々な硬さと密度のプレートを含み、観察できるイメージを検討してみた。皮膚や筋肉と大きく音響インピーダンスの異なる

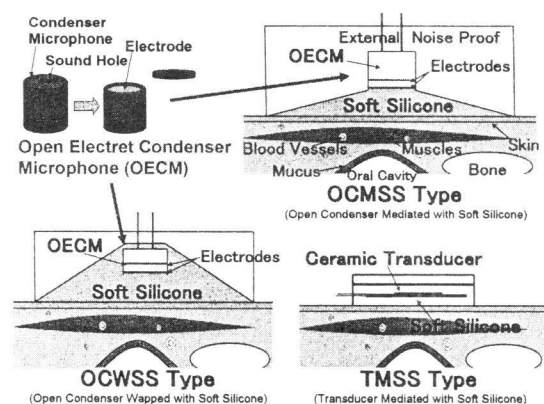


図 2 ソフトシリコン型 NAM マイクロホン  
Fig. 2 Soft silicone type NAM microphones.

金属を置いたときには、その境界面でほとんどが超音波が反射を起し、金属プレート以降の体内構造が全く描出されず、また前述のハードシリコンを用いた場合、微かに体内構造が描出された。人の軟部組織に近いと思われるソフトシリコンを用いると、何も挟まなかった場合と同様に、ほぼ完全に体内構造が描出された。つまり人間の軟部組織を伝導する音を可能な限り反射減衰させることなく OEMC の振動電極まで媒介するには、その軟部組織と同等の音響インピーダンスをもつソフトな物質を用いると効率の高いことが推察できる。

本来 OEMC を人体の軟部組織内に直接埋め込んだり、OEMC の振動電極を直接皮膚に密着させたりすることが可能ならば、それが体内伝導音をサンプリングする最上の手段である。しかしそれは現実的ではない。そこでソフトシリコンを皮膚に密着させ、皮膚に似た人工的な肉の突起物を作り、その中に OEMC を埋め込んだり、OEMC 振動電極を直接接触させたりすれば、上述の状態と音響インピーダンス的には同等の効果を作り出すことが可能であると考えた。

そこで人間の皮膚と同程度の柔らかさと弾性をもつソフトシリコンを音媒体に用いた NAM マイクロホン試作品を数多く作成したが、その構造や特性を分類すると、大きく次の三つのタイプに分類できる。

第 1 のタイプは、図 2 の右上のように聴診器型の空気部分に相当する円錐型の微小密閉空間の空気を、そのままソフトシリコンに置き換えた形である。

第 2 のタイプは図 2 の左下のごとくソフトシリコンで OEMC を完全に包埋してしまったタイプのものである。OEMC は電極面だけでなく側面や背面からの振動にも感度があるからである。またこの構造はマイク裏面や側背部から外部ノイズの浸透する領域と、皮膚表面から伝わる振動音の伝達する領域とを隔離可能であり、肉の中に OEMC を直接埋め込んだのと同様の効果が期待できる。

第 3 のタイプは、図 2 の右下のごとく円盤状のセラミック圧電素子（元来はプザー出力用途）と、音媒体としてソフトシリコンを組み合わせて用いるものである。完全に周囲を包埋して圧電素子をソフトシリコン内に浮かせるようなタイプのものから、圧電素子の辺縁や一部を固定して、片面だけをソフトシリコンで媒介したりするものなどがある。

以上 3 タイプとも、松風（株）製の歯科複模型用シリコン印象剤デュブリコン（DUPLICONE: vinyl polysiloxane）を用いた。

図 3 と図 4 に、上記三種類のソフトシリコン型 NAM マイクロホンでサンプリングした NAM 音と BTOS 音のスペクトルを、聴診器型と比較して呈示

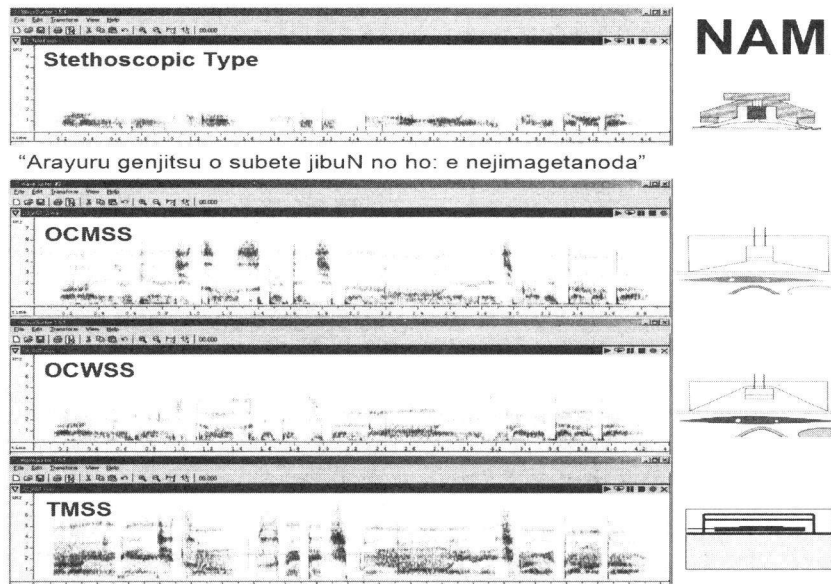


図 3 ソフトシリコン型 NAM マイクロホンによる NAM のスペクトル比較  
Fig.3 Comparison of NAM Spectra by soft silicone type NAM microphone.

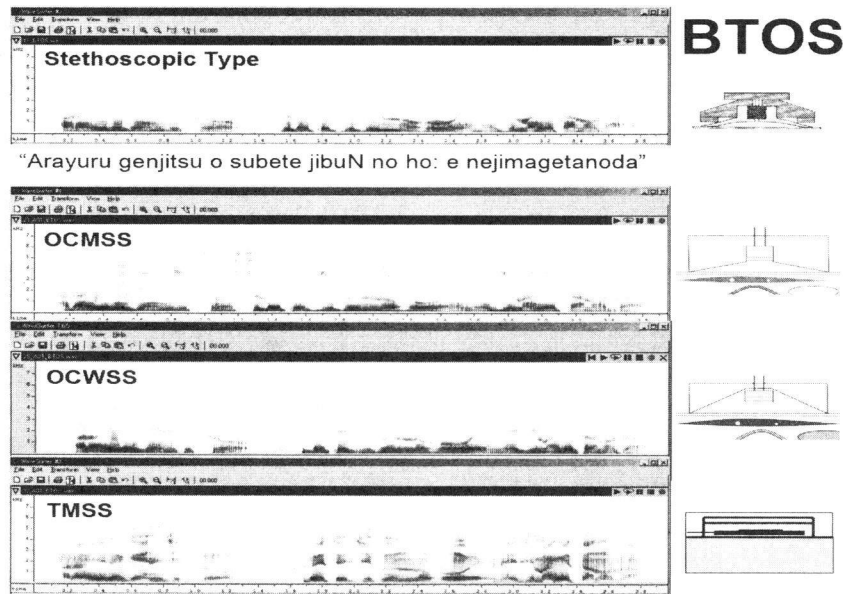


図4 ソフトシリコン型 NAM マイクロホンによる BTOS のスペクトル比較  
Fig.4 Comparison of BTOS spectra by soft silicone type NAM microphone.

する。発話内容は「あらゆる現実をすべて自分の方へねじ曲げたのだ」である。聴診器型では 2kHz 以上が全く描出されず第 2 フォルマントの一部が不明瞭となったが、ソフトシリコン型では、NAM においても BTOS においても、帯域が 4.5kHz から 7kHz まで伸び、はるかに高域のフォルマントまで描出される。

#### 4. 気導音声と肉伝導音声の同時ステレオ収録による比較

NAM マイクロホンには大きな二つの特性がある。一つは肉伝導音をどれくらい感度と帯域で拾えるかということ。二つ目は皮膚に装着して人間の頭と NAM マイクロホンをつなぐ大きな気導マイクロホンとみなしたときの気導音に対する感度の低さである。

NAM 音と、NAM 発話時に口から放射される「気導 NAM 発話音」とが、どの程度異なるのか、ソフトシリコン型 NAM マイクロホンと同型番の ECM を気導音マイクロホンとして用いて、同じ発話行為に対する気導音と肉伝導音の両方を同時に 2トラックでステレオ同期収録し、その両方の音声信号から、同じ発話様式を気導音の観点と肉伝導音の観点から客観的に比較する実験を行った。

NAM マイクロホンはタイプによって特性が違う。同じ設計に従って同じマイクロホンを製作しようとし

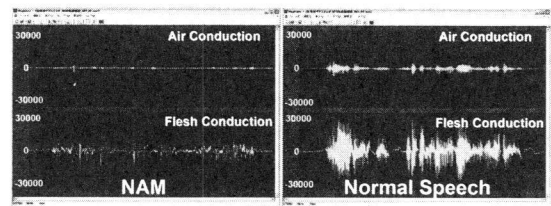


図5 ステレオ収録による気導音声と肉伝導音声の比較  
Fig.5 Comparison between air conduction and flesh conduction by stereo sampling.

ても、手作りのため特性にばらつきが出る。実験に使うソフトシリコン型 NAM マイクロホンとしては帯域、接触感度、雑音耐性について平均的な図 2 の OCMSS タイプのものを選んだ。

ステレオマイクアンプの気導音側と肉伝導音側の増幅率と出力レベルは同等にし、サンプリングレート 16kHz、量子化 16bit でステレオ収録した。気導音は NAM マイクロホンに使用したものと同一型番のコンデンサマイクロホンを、ヘッドセットマイクロホンのように、直接息がかからない程度に口元から横にずらし 5cm 離れた位置に固定して、室内静音環境で収録した。発話内容は、今までと同様に「あらゆる現実をすべて自分の方へねじ曲げたのだ」である。

図 5 の左は、気導 NAM 発話音（つまり NAM 発話

中に漏れ聞こえる呼吸音) と, NAM マイクロホンから収録される肉伝導音の信号の比較である. マイクアンプの出力レベルは 2トラックともに NAM の収録レベルに合わせた. 肉伝導音のパワーと気導音のパワーの, フレームごとの差の平均は, 5.98 dB であった.

図 5 の右は, 気導音としての通常音声と, NAM マイクロホンから収録される肉伝導の通常音声 BTOS である. マイクアンプの出力レベルは BTOS に合わせた. 肉伝導音のパワーと気導音の, フレームごとのパワーの差の平均は, 10.18 dB であった.

また気導音サンプリングと肉伝導音サンプリングにおける, 気導音環境雑音信号と目的音声信号の信号強度の違いを見るため, 実験者の前方 50 cm の距離のスピーカより TSP (Time Stretched Pulse) 音信号の繰返し音を環境雑音として出力しながら発話したのも収録した. 図 6 の左は, TSP 音信号を背景雑音として繰返し流しながら NAM 発話を行ったものである. マイクアンプの出力レベルは NAM に合わせた. 気導音の音声信号が微弱すぎて見にくい, TSP 音信号と音声信号との振幅の違いを上下 2トラックで比較すれば, 気導音環境と肉伝導音環境での目的信号と環境雑音信号の比率が分かる.

図 6 の右は, TSP 音信号を同音量で背景雑音として繰返し流しながら通常音声発話を行ったものである. マイクアンプの出力レベルは肉伝導音の BTOS に合わせた. NAM に比べてゲインを通常 20 dB 程度下げることができる. この肉伝導の BTOS の場合, TSP 音信号が脈音と首の筋振動による基線の揺れの中に隠れて見えなくなってしまう.

同型番マイクロホンによる同期ステレオ収録にて, 同じ発話様式の, 気導音としての信号と, ソフトシリコーン NAM マイクロホンに加工して肉伝導音としてとらえた信号とを比較した. NAM でも BTOS で

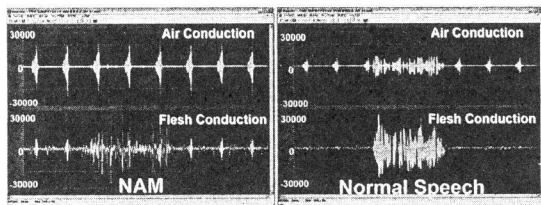


図 6 ステレオ収録による気導音声と肉伝導音声の比較 (TSP 音を背景雑音として繰返し流したもの)

Fig. 6 Comparison between air conduction and flesh conduction by stereo sampling with iterative TSP sound on the background.

も, 目的信号である音声の感度はソフトシリコーン型 NAM マイクロホンを使用した肉伝導音の方が, 気導音より高い. また気導音の環境雑音信号は肉伝導音で低くなるのが分かる. 特に BTOS の場合の対環境雑音耐性は顕著である.

### 5. 話者適応による認識実験

聴診器型 NAM マイクロホンに比した, ソフトシリコーン伝導型 NAM マイクロホンの数値的な評価として, 話者適応 (Iterative MLLR) による NAM 音響モデルと BTOS 音響モデルの作成を行い, HMM による認識率を比較した.

新聞記事 400 文と, 評価用の新聞記事 24 文 [6] を, 特定男性話者が NAM 発話により同じ NAM マイクロホンで読み上げて 16 kHz サンプリングした. 通常音声男性不特定話者の Phonetic Tied Mixture (PTM) モデル (64 混合, 3000 状態) [6] を初期モデルとして HTK [18] を使い, 400 文章, 128 クラスタで 10 回の繰返し話者適応 (Iterative MLLR) [17] を行った. 認識エンジンは Julius3.4.2 [6] を使い, 言語モデルの辞書として 20K 辞書 [6] を用いた. 認識率の評価は JDTK [6] を使い, 単語認識精度を計算した.

図 7 はソフトシリコーン型 NAM マイクロホン 3 種でサンプリングした NAM と, 聴診器型 NAM マイクロホンでサンプリングした NAM, そして対照としてヘッドセットマイクロホンで NAM とは別に収録したささやき声の, Iterative MLLR による認識率の比較を示すグラフである (横軸は MLLR の繰返し回数). OCMSS 型, OCWSS 型, TMSS 型の 3 タイプ

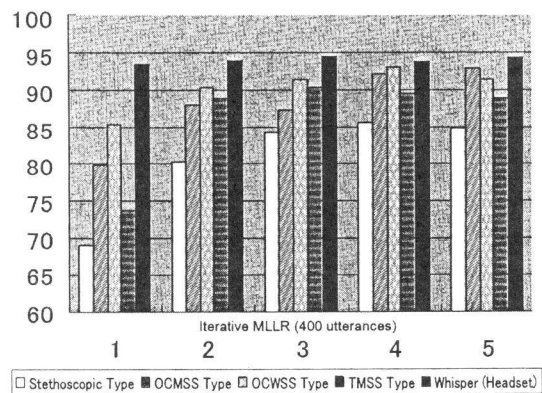


図 7 Iterative MLLR による NAM 認識の単語認識精度  
Fig. 7 NAM recognition accuracy by iterative MLLR.

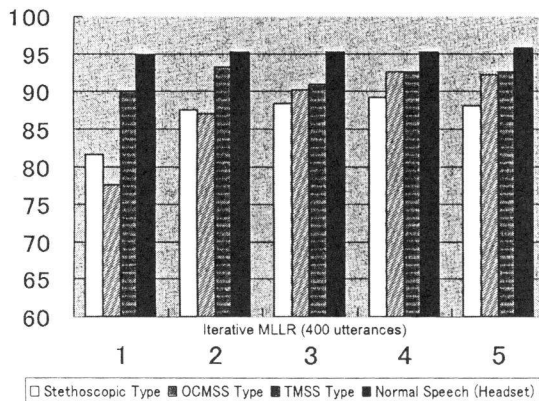


図8 Iterative MLLRによるBTOS認識の単語認識精度  
Fig. 8 NAM recognition accuracy by iterative MLLR.

ともにソフトシリコン型 NAM マイクロホンは聴診器型 NAM マイクロホンの単語認識精度を上回った。

図8はソフトシリコン型 NAM マイクロホン2種でサンプリングしたBTOSと、聴診器型 NAM マイクロホンでサンプリングしたBTOS、そして対照としてヘッドセットマイクロホンで収録した通常音声の、Iterative MLLRによる認識率の比較を示すグラフである。ここでもOCMSS型、TMSS型の2タイプのソフトシリコン型 NAM マイクロホンは聴診器型の単語認識精度を上回った。OCMSS型はBTOSの場合、同じマイクアンプでは感度が高すぎて、収録文章のほぼすべてがオーバーフローによる音割れを起こしていたため、この比較のグラフには含めなかった。

図7でも図8でも対照(黒い棒グラフ)の気導音声と比べると、肉伝導音声の単語認識精度は低い。これが気導音声に比べて肉伝導音声の情報が不足しているためであるのか、これらの話者適応がすべて初期モデルとして気導通常音声の不特定話者モデルを用いているためであるのか、またその両方であるのかは不明である。NAMのみのサンプルから作られた不特定話者モデル、BTOSのみのサンプルから作られた不特定話者モデルの作成が大きな課題となる。

## 6. 肉伝導音声の聞き取り実験

聴診器型 NAM マイクロホンを用いてサンプリングしたNAMも、初めて聞いて分かる場合もあるし、何度も聞いているうちに耳に慣れが生じて、かなり内容が分かるようになるものである。新しいソフトシリコン伝導型 NAM マイクロホンでは、より高域に

帯域が伸びて音質が改善され、聴診器型に比し、より聞き取りやすくなった。音源付与や声質変換などを行い、無音声電話の入力にNAMを利用するにしても、NAMやBTOSそのものを通信に使用するにしても、サンプリングされたオリジナルのNAM音やBTOS音に人間が聞き取りを行うために有効な情報が可能な限り多く入っていることが望ましい。それはとりもなおさず信号処理前のNAM音そのものを様々な人に聞き取ってもらって、その聞き取り精度が高いほど良いことになる。HMMを用いた機械認識では、原理的にいって音響モデルの作り方を考えれば、必ずしも人が長年聞き慣れた気導音声に近いほど認識率が良いということにはならない。しかし少なくとも最終出力を人の耳で聞く場合は、気導音声に近い方が聞き取りやすい。そこでソフトシリコン伝導型 NAM マイクロホンの数値的評価のもう一つの側面として、実際に人によるNAM音声とBTOS音声の聞き取り実験を行った。

### 6.1 聞き取り実験の方法

10数単語程度の文章(日常的な用件文が主体)を12個、意味の分かる単独の単語のみを12個、そして全く意味をなさない3~4モーラ程度の無意味な擬似単語12個を用意した。文章と意味単語は中学生程度の語彙で十分理解できる程度のものに限定し、知名度の高い言葉でも特定の分野に興味がないと分からない単語は排除した。つまり一般人の語彙知識に必ず存在すると思われる言葉を選んだ。未知語の認識、つまり音素単位の認識は無意味単語で評価した。

以上の36問題を、男性話者1名が標準的な速度のNAM発話とBTOS発話で読み上げたサンプルを、新旧2種類のNAMマイクロホンで録音し、対照とするため気導音の通常音声やささき声もヘッドセットマイクロホンにて録音した。すべてについて8kHzと16kHzの2種類のサンプリングレートでデジタル化し(8kHzは16kHzからダウンサンプリングして作成)、各聞き取り問題ごとに12種類の録音を作成した。つまり3種類のマイク×2種類の発声法×2種類のサンプリングレート=12種類の録音である。なおソフトシリコン伝導型 NAM マイクロホンとしては、聴診器型と比較する上でその構造的に類似点が多いが、音媒体の全く異なる前述のOCMSSタイプのものを用いた。音量は、実際に聴取したり音声信号の最大振幅を見て、大きく変わらない程度に配慮した。

NAMやBTOSの聞き慣れを防ぐため、上記12種

類の録音をランダムに混在させることにした。被験者12名用に12個のテストセットを作成した。それぞれのテストセットでは同じ文章や単語を同じ順に聴取させるが、テストセット中の $12 \times 3 = 36$ 個のそれぞれの問題については12種類の録音の被験者への割当方を無作為の順列とした。各問題につき12種類の録音をすべて登場させて、しかもだれに割り当てるかを無作為にする。これで各人は上記12種類の録音をランダムに聞くことにもなる。しかも最終的に使用されない録音サンプルはなく、12種類の録音のトータルの出現数は同数となる。被験者として、NAMやBTOSを聞き慣れていない一般の10代から60代までの男性6名、女性6名の計12名を選出した。各世代ごとに男女各1名ずつを原則に選び、平均年齢は男41.83歳、女38.83歳であった。これをまたクジでランダムに12のテストセットに割り振った一つの問題について同じ録音を全部3回ずつ密閉式ヘッドホンで聴取し、聞いた結果をそのたびごとに答案用紙に書き取らせた。全員が $12 \times 3 \times 3 = 108$ 個の録音を聴くことになる。認識率の計算方法は、音声認識の場合と同様であり、評価に単語認識精度を用いた。

## 6.2 実験結果

まず各発声法と各マイクロホン別に、サンプリングレートで単語認識精度に差が見られるかを検討した。使用したのは最初の12個の文章聞き取りの結果である。各収録方法においてサンプリングレートが16kHzであったか8kHzであったかということで、単語認識精度に有意な差は見られなかった。対照のヘッドセットマイクロホンで収録したささやき声と通常音声についても結果は同じであった。そこでサンプル数を増やして認識精度の統計的信頼性を増すため、両者を混合して、文章の単語認識精度を計算した。

図9はNAMの文章聞き取りの単語認識精度である。対照としてヘッドセットマイクロホンで収録したささやき声（明らかに気導音としての音量は気導NAM発声音より大きい）を対比して掲げる。

NAMの文章聞き取りについては、旧式の聴診器型NAMマイクロホンに比して、ソフトシリコーン伝導型NAMマイクロホンは、どの聞き取り回数においても有意に単語認識精度が高く、対照である気導音のささやき声の単語認識精度と2回目、3回目の聞き取りについては有意差がなかった（95%信頼区間）。1回目の認識率の低さは、ささやき声が普段聞き慣れているのに対し、NAMはほぼ初めて聞く人ばかりであった

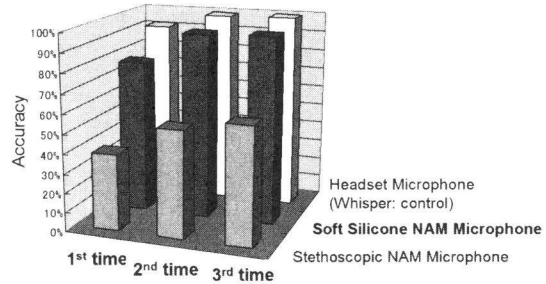


図9 NAMの文章聞き取り単語認識精度  
Fig.9 Hearing accuracy of sentence (NAM).

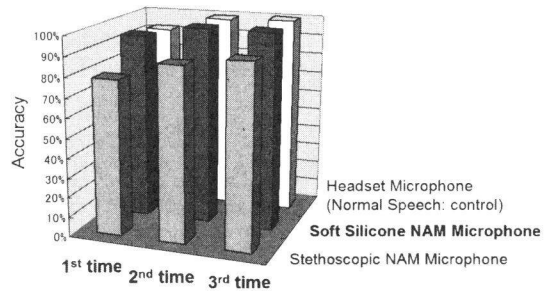


図10 BTOSの文章聞き取り単語認識精度  
Fig.10 Hearing accuracy of sentence (BTOS).

ためと考えられる。また聞き取れているにもかかわらず文章の部分的聞き忘れが1回目にはあるが、2回目以降は1回目の記述を踏まえて聞くことができるため、聴覚をより反映すると考える。

図10はBTOSの文章聞き取りの単語認識精度である。これに見られるように、図9と比較して、有声音子音、無声音子音の鑑別が容易であるためか、各収録法でBTOSの方がNAMより高い単語認識精度を示す。BTOSでもソフトシリコーン伝導型NAMマイクロホンは、対照であるヘッドセットマイクロホン収録通常音声の単語認識精度と有意差はなく、聴診器型NAMマイクロホンに比べて1回目、2回目、3回目の聞き取りで有意に高かった。

文章の場合は、機械認識と異なり、人間は言語モデル以上の語彙知識をもつ以外にも、文脈から文章を類推することが可能である。しかし単独単語認識では、語彙知識こそ使えるが文脈がなく、より高次の知識でこれを補うことができない。また無意味単語では、語彙知識は通用せず、全く音素の聞き取りとなる。一つでも音素を聞き違えると、置換誤りとして単語認識精度を計算した。そのためこのテストが最も厳しい条件のテストとなっている。図11と図12にNAMと

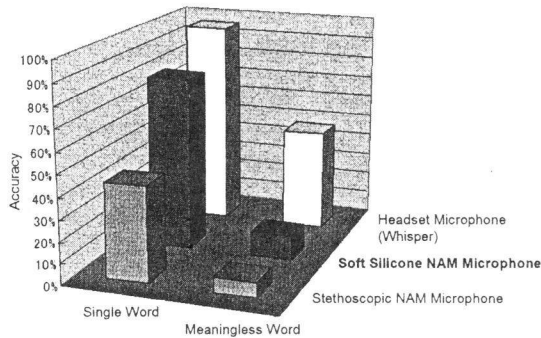


図 11 NAM の単独単語聞き取り精度  
Fig. 11 Hearing accuracy of single word (NAM).

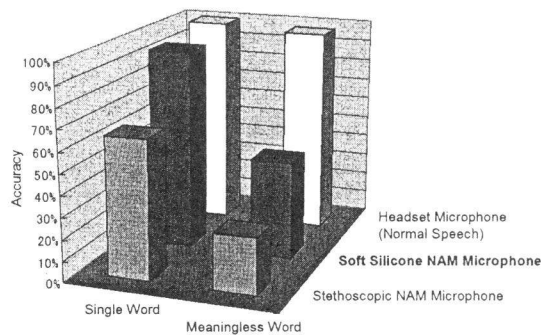


図 12 BTOS の単独単語聞き取り認識精度  
Fig. 12 Hearing accuracy of single word (BTOS).

BTOS の単独単語の単語認識精度を示す。意味単語が左側の棒グラフの列、無意味単語が右側の棒グラフの列である。また意味単語の場合も、無意味単語場合も 8kHz サンプリングと 16kHz サンプリングにはやはり有意な差が見られなかったため両者をまとめて計算した。文章と異なり、単語の場合は 3 回の繰り返し聴取による有意な改善は見られなかったため、3 回の平均を掲示した。

意味単語においては NAM でも BTOS でも、図 11 と図 12 の左の列のグラフに見られるとおりの文章の聞き取りと同様の傾向を示し、ソフトシリコン型は聴診器型に比し気導音声にかなり接近する。

無意味単語の場合は、図 11 と図 12 の右の列グラフに見られるとおりのソフトシリコン型 NAM マイクロホンでも極端にその認識精度が低くなり、図 11 の左の列と比較すると分かるように、特に NAM において低下が著しい。とはいえ無意味単語では、気導通常音声で 90%以上の認識精度を示すのに比し、もともと気導音声のささやき声でさえ、認識精度は 50%を切

る。これは有声子音と無声子音の判別などが難しいためであろうと推察される。気導音声でも無意味単語に対しては「有声音主体の通常音声」と「無声音のささやき声」にこれほど大きな聞き取り認識率の違いがある。これは文章や意味単語の場合ならば、人間の知識と知能がこれを補うが、これの通用しない音素の聞き取りは、ささやき声や肉伝導音声の場合完全ではないからであると考えられる。

### 6.3 聞き取り実験のまとめ

ソフトシリコン伝導型 NAM マイクロホンは、人間の聞き取りにおいて認識の精度を、旧式の聴診器型よりも大きく高めた。しかし無意味単語のような音素そのものを聴き取らなければならない状況で、認識率は特に NAM において気導音声に大きく劣る。電話などでの会話中、定型文を話しているうちはいいが、未知語が話題に登場したときに会話を成立させるためには、この音素の判別性を上げなければならない。またこれは今後の NAM 音の信号処理による通常音声化の研究や BTOS の信号処理による音質変化の研究に多いに期待される点である。

## 7. む す び

ソフトシリコンを音媒体とする新しいタイプの NAM マイクロホンは、気導音マイクロホンに比して、環境雑音信号を抑えつつ、目的音声信号をより大きく収録できる。上限帯域 2kHz だった聴診器型に比し、より広い帯域をもち、肉伝導音声の HMM 認識率を向上させたばかりではなく、未信号処理のまま聴取しても聞き取りやすくなった。これにより NAM 音自身や、またはそれに声質変換や音源付与などの信号処理を加え通常音声化したものを直接通信に用いることが可能であり、環境の影響を受けにくく周囲の状況に気兼ねしない、いわゆる「無音声電話」などへ応用が考えられる。また接触面感度や外部雑音耐性においてもより優れたものが作成可能である。加えて可塑性をもつ素材を選択したことにより、マイクロホン形状と大きさがより自由に設計でき、小型化・薄型化が可能になった。今後体表への固定法の改良や無線化も含めて、NAM を機械認識だけではなく汎用音声入力インタフェースとする非環境依存型コミュニケーションのための、使いやすい日常使用型 NAM マイクロホンデザインが課題である。

## 文 献

- [1] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari,



- and K. Shikano, "Accurate hidden Markov models for non-audible murmur (NAM) recognition based on iterative supervised adaptation," Proc. ASRU, pp.171-185, 2003.
- [2] P. Heracleous, Y. Nakajima, A. Lee, Hiroshi H. Saruwatari, and K. Shikano, "Audible (normal) speech and inaudible murmur recognition using NAM microphone," Proc. 12th European Signal Processing Conference (EUSIPCO2004), pp.329-332, Sept. 2004.
- [3] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, and K. Shikano, "Non-audible murmur (NAM) speech recognition using a stethoscopic NAM microphone," Proc. 8th International Conference on Spoken Language Processing (ICSLP2004), WeC2102p-6, pp.527-530, Oct. 2004.
- [4] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, and H. Shikano, "Non-audible murmur (NAM) recognition exploiting adaptation techniques," IEICE Technical Report, SP2003-170, Jan. 2004.
- [5] S.C. Jou, T. Schultz, and A. Waibel, "Adaptation for soft whisper recognition using a throat microphone," Proc. 8th International Conferences on Spoken Language Processing (ICSLP2004), WeC2102p.12, Oct. 2004.
- [6] 河原達也, 李 晃伸, 小林哲則, 武田一哉, 峯松信明, 嵯峨山茂樹, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田 篤, 宇津呂武仁, 鹿野清宏, "日本語ディクテーション基本ソフトウェア (99年度版)," 音響誌, vol.57, no.3, pp.210-214, 2001.
- [7] 李 晃伸, 河原達也, 鹿野清宏, "既述文法に基づく高性能連続音声認識エンジン Julian," 音響講義集, 3-1-10, pp.111-112, Oct. 2001.
- [8] 中島淑貴, 柏岡秀紀, ニック キャンベル, 鹿野清宏, "非可聴つぶやき認識," 信学論 (D-II), vol.87-D-II, no.9, pp.1757-1764, Sept. 2004.
- [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] 中島淑貴, 竹苗浩司, 柏岡秀紀, 鹿野清宏, ニック キャンベル, "NAM interface communication," 情処学研報, SLP-52, vol.2004, no.74, pp.33-40, July 2004.
- [12] 中島淑貴, 柏岡秀紀, 鹿野清宏, ニック キャンベル, "微弱体内伝導音抽出による無音声認識," 音響講義集, 3-Q-12, pp.175-176, March 2003.
- [13] 中島淑貴, 柏岡秀紀, 鹿野清宏, ニック キャンベル, "無音声認識におけるセンシング方法の改善," 日本音響学会講演論文集, 3-Q-1, pp.145-146, March 2004.
- [14] 竹原靖明, 腹部エコーの ABC (日本医師会生涯教育シリーズ), 医学書院, 東京, 1991.
- [15] 竹苗浩司, 中島淑貴, Panikos Heracleous, 李 晃伸, 猿渡 洋, 鹿野清宏, "雑音下における Non-Audible Murmur (NAM) 認識の頑健性の評価," 音響講義集, 1-1-16, pp.31-32, Sept. 2004.
- [16] 戸田智基, 鹿野清宏, "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," 情処学研報, SLP-54, vol.2004, no.74, pp.157-162, 2004.
- [17] P.C. Woodland, D. Pye, and M.J.F. Gales, "Iterative unsupervised adaptation using maximum likelihood linear regression," Proc. ICSLP, pp.1133-1136, 1996.
- [18] S. Yong, G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, "The HTK book (for HTK Version 3.2.1)," Cambridge University Engineering Department, 2002.

(平成 17 年 6 月 1 日受付, 11 月 1 日再受付)

#### 中島 淑貴 (正員)



昭 62 東大・医・医卒。内科基礎研修を経て平元～3 帝京大学医学部法医学教室に助手として勤務。平 3～7 消化管内視鏡を中心に、内科医として民間病院にて臨床に従事。平 7～平 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。専門は「吐血」。平 17 奈良先端大情報科学研究科後期博士課程了。平 17 ATR 人間情報科学研究所・生体イメージング研究室客員研究員。平 18 奈良先端科学技術大学院大学情報科学研究科勤務。平 15 ISID 学生論文大賞受賞, 平 15 日本音響学会ボスター賞, 平 17 本会論文賞・猪瀬賞受賞。日本医師会認定産業医, 日本内科学会, 消化器内視鏡学会, ISCA, 音響学会各会員。

#### 鹿野 清宏 (正員:フェロー)



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, 平 2 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, 平 17 本会論文賞・猪瀬賞。IEEE, ISCA, 情報処理学会, 音響学会, VR 学会各会員。