2004－ＨＩ－109　(6)
2004－SLP－ 52　(6)
2004／7／16

NAM Interface Communication

中島淑貴　竹苗浩司　柏岡秀紀　鹿野清宏　ニックキャンベル

要旨：我々は非可聴の調音呼吸音である非可聴つぶやき(Non-Audible Murmur: NAM) を定義し，これを耳介後下方部の皮膚に接した自作 NAM マイクロフォンよりサンプリングして HMM による大語彙連続認識が可能であることを示した．NAM 音は認識だけではなく，モーフィングや音源付与などによる通常音声化が可能であり，無音声電話などへの応用が期待されている．そのためには従来の聴診器型 NAM マイクロフォンでサンプリングされる NAM は 2KHz 以上にフォルマントがほとんど見られず，その帯域の広範囲化と感度の向上などの音質向上が不可欠である．今回我々は皮膚の音響インピーダンスに近いソフトシリコーンを音媒体に用いて NAM マイクロフォンの抜本的な改良を行い，その帯域の広範化と接触面感度の上昇を得て，NAM 認識の認識率と音質が飛躍的に向上したので報告する．

NAM Interface Communication

Yoshitaka Nakajima, Koji Takenae, Hideki Kashioka, Kiyohiro Shikano, Nick Campbell

SUMMARY: We are working towards the next generation of skin-attachment sensors for sampling NAM signals (Non-Audible Murmur) by using soft silicone, which has an acoustic impedance close to that of human flesh, as the prime medium of vibration. These new NAM microphones enable us to record more wideband NAM sound, which is clear enough to convey articulated utterances even without digital signal processing, and provides increased robustness against noise in comparison with the former stethoscopic type. We obtain a much higher accuracy of NAM recognition and, as a result, suggest the possibility of a “Non-Voice Phone”.

1. はじめに

非可聴つぶやき（Non-Audible Murmur: NAM）という言葉は,「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．

我々は自作したセンサー（NAM マイクロフォン）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して NAM 音信号を多数サンプリングし，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，認識エンジン Julius を用いて NAM による大語彙連続認識が可能であることを示した[1, 2, 3, 4].

NAM 認識による，いわゆる無音声認識の可能性に加えて，NAM 音は増幅して信号処理を施さずに聞いても,「音のこもったささやき声」としてある程度聞き取り可能なため，声質変換や音源付与などの信号処理により，直接通信に使う事も可能と考えられる．これが実現すれば，いわゆる無音声電話等の応用が考えられ，先の無音声認識と合わせて，図 1 のように，人対人，人対機械のコミュニケーションに幅広く NAM がインターフェースとして使用できる．これは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，極めてユニバーサルなデザインの新しいインターフェースとなる．

* 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

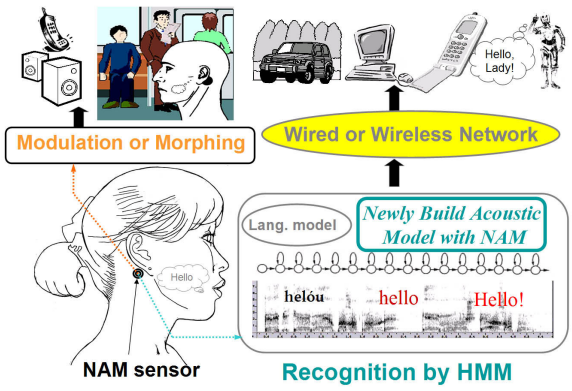


Figure1: NAM Interface Communication

それを実現するに当たって，認識においても，直接通信においても，サンプリングされるオリジナルの NAM 音信号の音質を可能な限り向上させて判別性を上げる事が，重要課題である．今回我々は NAM マイクロフォンの抜本的な改良を行った．

2. 聴診器型 NAM マイクロフォン

NAM による大語彙連続認識が可能であることを示した自作 NAM マイクロフォンは図 2 の如く医療用聴診器の原理を応用したものであった．内蔵されたコンデンサマイクロフォン (Electret Condenser Microphone: 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．

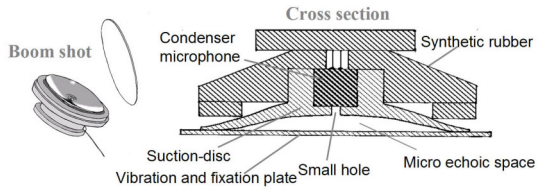


Figure2: Stethoscopic NAM Microphone

NAM マイクロフォンでゲインを調節して収録した通常音声を体内伝導通常音声（Body Transmitted Ordinary Speech: BTOS）と定義するが，この聴診器型 NAM マイクロフォンでは，NAM 信号についても BTOS 信号についても，2KHz 以下にしかスペクトラムにフォルマントが描出されない．図 3 の上段は通常の気導マイクロフォンで 16KHz サンプリングしたささやき声（左）と通常音声（右）を比較のために掲げる．下段は聴診器型 NAM マイクロフォンによる NAM 音（左）と BTOS 音（右）のスペクトラムである．

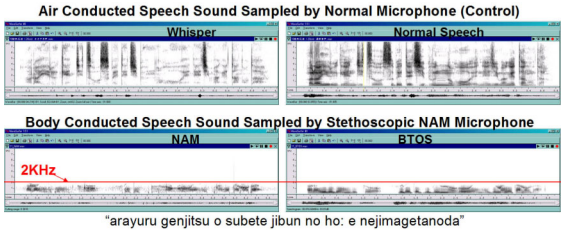


Figure 3: NAM and BTOS signals by stethoscopic NAM Microphone in Comparison with Air Conducted Voices

NAM 音を増幅してそのまま聞いた印象は「こもったささやき声」に近く，会話内容を聴取することは困難であり，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，判別性は当然 NAM より良いが，「非常に音のこもった通常音声」となる．

3. シリコーン型 NAM マイクロフォン

3.1 帯域を広範化させるために

聴診器型 NAM マイクロフォンにおける三角錐形状の微小密閉反響空間は，低域の感度を上昇させる利点もある反面，この空気層がむしろ帯域を狭小化させているのではないかと考えられた．そこで振動板と ECM との間に介在する音媒体としての空気を排することを試み，新たに音媒体として，弾性があり形成が容易で安全な歯科技工用のシリコーンを選択した．

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 表面に開けた小孔を通じて空気の振動を振動電極版に伝える構造となっている．我々は図 4 の左の如く，この小孔の開いた 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．これを仮に Open Electret Condenser Microphone (OECM)と名付ける．

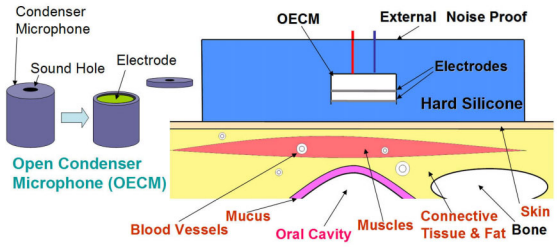


*Figure 4: Open Condenser Wrapped with Hard Silicone Type NAM Microphone (**OCWHS**)*

この OECM を，図 4 の右のように硬質の消しゴム程度の硬さのハードシリコーンに完全に包埋し，接着剤を使用せず直接皮膚に圧着してみた．これを Open Condenser Wrapped with Hard Silicone (OCWHS) 型の NAM マイクロフォンと呼ぶことにする．これを用いると，聴診器型に比し，感度はやや落ちるものの，図 5 の NAM 音のスペクトラムに見られる如く 2KHz 以上の帯域もやや表現されるようになり 2～3KHz のフォルマントも明瞭となった．

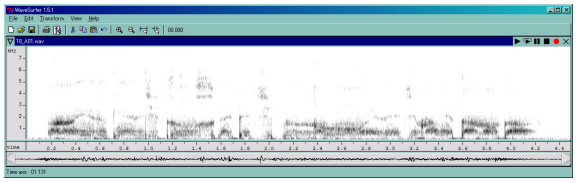
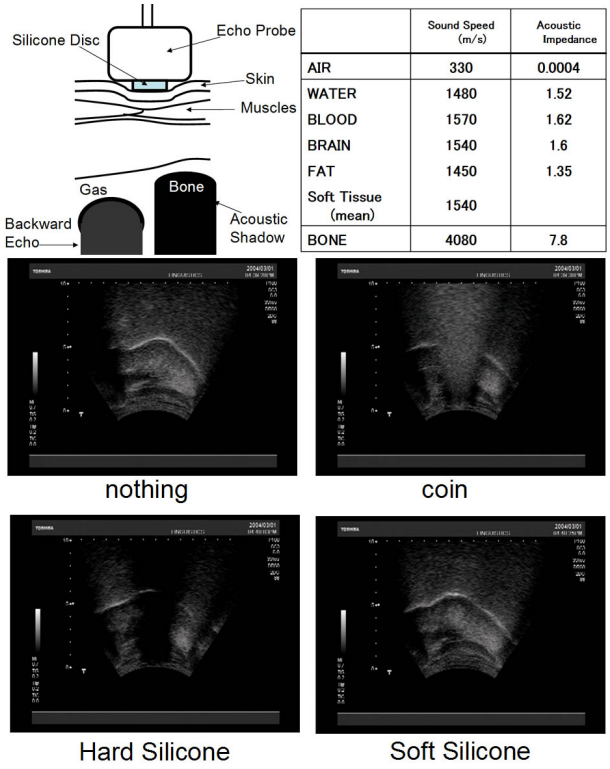


Figure 5: Spectrum of new NAM sampled with OCWHS

3.2 接触面感度を上昇させるために

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，音のこの性質を利用している．

図 6 の左上のように，超音波プローブと皮膚との間にさまざまな硬さと密度のプレートをはさみ，観察できるイメージを検討してみた．図 6 の 4 枚の写真に見られる様に，皮膚や筋肉と大きく音響インピーダンスの異なる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こし，金属プレート以降の構造が全く描出されず，また前述のハードシリコーンを用いた場合，微かに高輝度のラインのみ描出された．人の軟部組織に近いと感じられるソフトシリコーンを用いると，何もはさまなかった時と同様にほぼ完全に構造が描出された．つまり人間の軟部組織を伝搬する音を可能な限り反射減衰させることなく OECM 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される．



	Sound Speed (m/s)	Acoustic Impedance
AIR	330	0.0004
WATER	1480	1.52
BLOOD	1570	1.62
BRAIN	1540	1.6
FAT	1450	1.35
Soft Tissue (mean)	1540	
BONE	4080	7.8

Figure 6: Visible Difference of Acoustic Impedances using Medical Ultrasonography

3.3 ソフトシリコーン NAM マイクロフォン

そこでソフトシリコーンを音媒体に用いた NAM マイクロフォン試作品を数多く作成してみたが，その構造や特性を分類すると，大きく三つのタイプに分類できる．もともと OECM を人体の軟部組織内に直接埋め込んだり，OECM 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ベストかも知れない．しかしソフトシリコーンを皮膚に密着させ，皮膚に人工的な肉のコブを作って，その中に OECM を埋め込んだり，OECM 振動電極を直接接触させれば，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できる．

第一のタイプは，図 7 のように聴診器型の空気部分に相当する三角錐型の微小密閉空間の空気を，そのままソフトシリコーンに置き換えた形である．

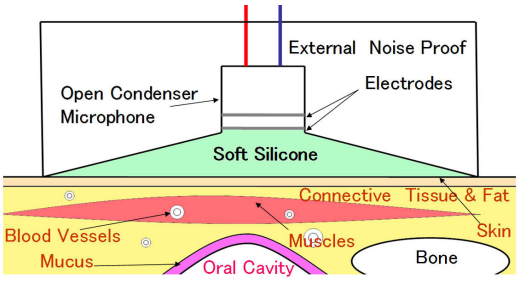


*Figure 7: Open Condenser Mediated with Soft Silicone Type NAM Microphone (**OCMSS**)*

第二のタイプは図 8 のようにソフトシリコーンで OECM を完全に包埋してしまったタイプのものである．OECM は電極面だけでなく側面や背面からの振動もかなり多く拾うからであり，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が隔離しやすい．ちょうど肉の中に OECM を直接埋め込んだのと同様の効果が期待できる．

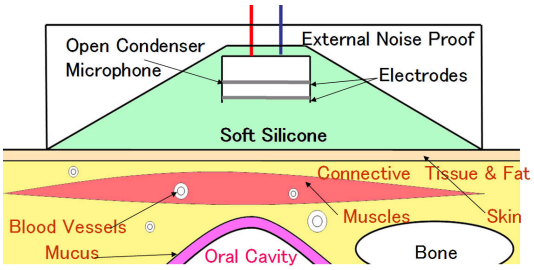


*Figure 8: Open Condenser Wrapped with Soft Silicone Type NAM Microphone (**OCWSS**)*

第三のタイプは，図 9 のように円盤状のセラミック圧電素子（元来はブザー出力用途）と，音媒体としてソフトシリコー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ー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ーンで媒介したりするものなどがある．

以上 3 タイプとも，シリコーン素材として，松風（株）製の歯科複模型用シリコーン印象剤デュプリコーン(DUPLICONE: vinyl polysiloxane)を用いた．

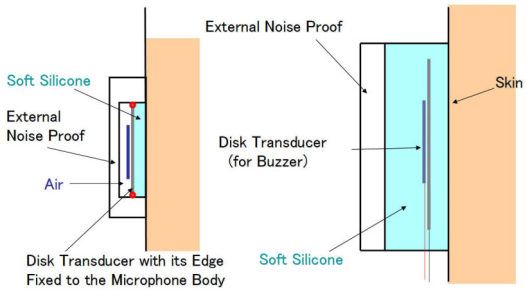


*Figure 9: Transducer Mediated with Soft Silicone Type NAM Microphone (**TMSS**)*

4. NAM マイクロフォンの簡易評価法

4.1 帯域

帯域の簡易評価として，図 2 のように NAM マイクロフォンで 16KHz サンプリングされた NAM 音や BTOS 音のスペクトラムを，通常気導マイクでサンプリングされた通常音声やささやき声のスペクトラムと比較する．

4.2 接触面感度

NAM マイクロフォンの接触面感度は,マイクアンプの入力ゲインボリュームを一定にしたときの NAM 信号の最大振幅で評価することとする．

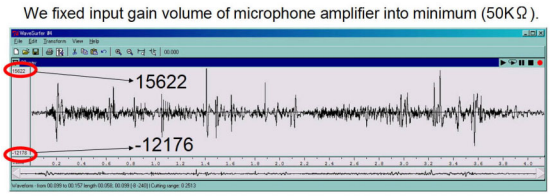


Figure 10: Contact Sensitivity

4.3 外部雑音への耐性

NAM マイクロフォンは肉の振動をサンプリングするために設計されている．肉のフィルタを一度通すわけであるから，当然外部雑音のレベルは低下するが，気導音をサンプリングする通常マイクロフォンとは異なった評価法があってしかるべきである．そこで NAM マイクロフォン単体での接触面感度とは別に，「NAM マイクロフォンを人間の頭に装着した状態」を一つの大きな仮想マイクロフォンとみな

すことを提案する．これを NAM microphone with human filter (NMHF) と呼ぶことにする．また外部雑音はあまりにも多くの種類があるため，外部雑音源として一定距離からの Transit Signal Priority (TSP) 信号の繰り返しを用いて，NMHF の周波数応答を測定してみた．マイクアンプの入力ゲインは認識や聴取に理想的な NAM 音や BTOS 音の振幅が得られるように，それぞれの NAM マイクロフォンのセンサー部に応じて調節する．この NMHF の TSP に対する応答が低ければ低いほど，つまり NMHF という仮想マイクの感度が悪ければ悪いほど，外部雑音に対して頑強であるあることになり，またその応答の曲線を見れば，どの周波数帯域の雑音に強いか弱いかがわかる．

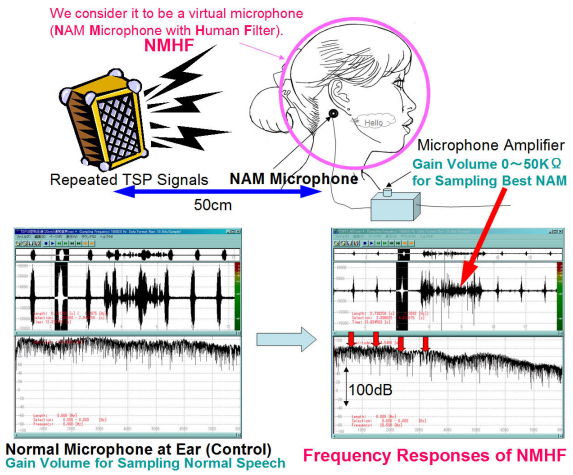


Figure 11: NMHF and Robustness against Noise

図 11 がこの評価法の概念図である．下段の左の図は，対照として，繰り返し大音量で TSP 信号が放出されるスピーカーから約 50cm の距離で耳元に置いた通常の気導マイクからサンプリングしたときの応答である．入力ゲインは通常音声が適度に収録できるように調節する．右図は同環境下で NAM マイクロフォンを NAM 音が理想的にサンプリングできるようにゲイン調節したときの NMHF の応答である．NAM マイクロフォンには，接触面から漏入する体内伝導の外部雑音と，マイクの側背部から漏入する外部雑音が混入するが，NAM や BTOS を最適音量でサンプリングしたときの外部雑音は大きく低減することが見て取れる．

4.4 タイプ別 NAM マイクロフォンの簡易評価

この簡易評価法で，聴診器型を対照として，ソフトシリコーン型 NAM マイクロフォンの三つのタイプに属する代表的な NAM マイクロフォンを選び，前述の如くその特性の簡易評価を比較してみた（図 12～15）．収録はある男性の特定話者より，同じマイクアンプを入力ゲイン調節して使用し，NAM 信号と BTOS 信号を 16KHz サンプリングした．外部雑音源としての繰り返し TSP 信号はすべて同じ音量で，同じ距離に頭を置いている．

大語彙連続認識が可能であることを示した旧型の聴診器タイプでは，医療用聴診器に劣らぬ感度を見せるが，NAM においても BTOS においても，前述の如く 2KH z 以上に急激なカットオフを認める．またこのモデルは 1KH z 周辺の低域の外部雑音に対し弱いことがわかる（図 12）．

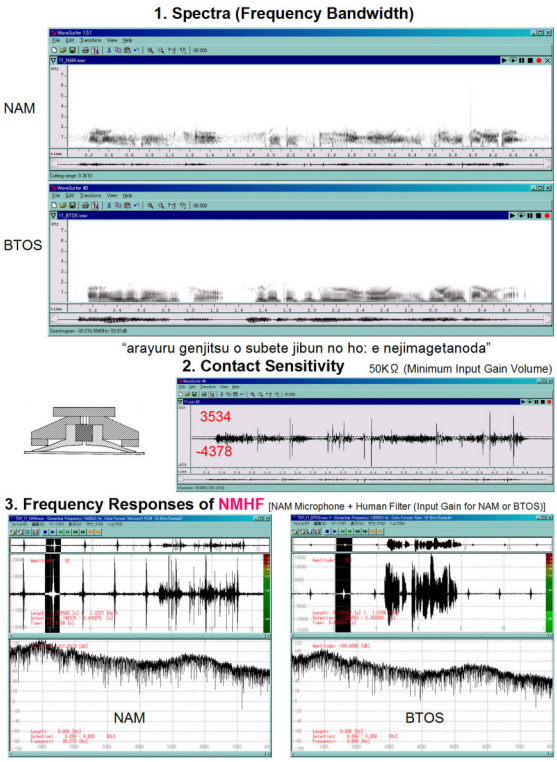


Figure 12: Stethoscopic Type

OCMSS タイプでは図 12 に見られるように, 聴診器型に比しはるかに広帯域の NAM 信号と BTOS 信号が得られ，しかも信号振幅の比較でわかるように接触面感度が高くなる．また NAM 信号においても BTOS 信号においても信号処理を加えなくとも増幅して聞けば, 発話内容はほぼ完全に聴取可能である.

OCWSS タイプは，図 14 の 2 の如く 3 タイプのうち最も高い接触面感度を示す．そのため入力ゲインを大幅に絞ることが可能で，外部雑音に対し最も頑強である. 帯域は OCMSS タイプに比べやや狭い.

TMSS は図 15 に示すように最も帯域が広く低域も強調されず，増幅して聞いたときの印象もささやき声や通常音声に最も近く明瞭である，しかし接触面感度は最も低かった.

一般に BTOS 信号を適音量でサンプリングする場合にはマイクアンプの入力ゲインを大きく下げることが可能なため，NMHF の感度は極端に低く，外部雑音としての大音量 TSP 信号もほとんど聞き取れないくらい低下させることができる.

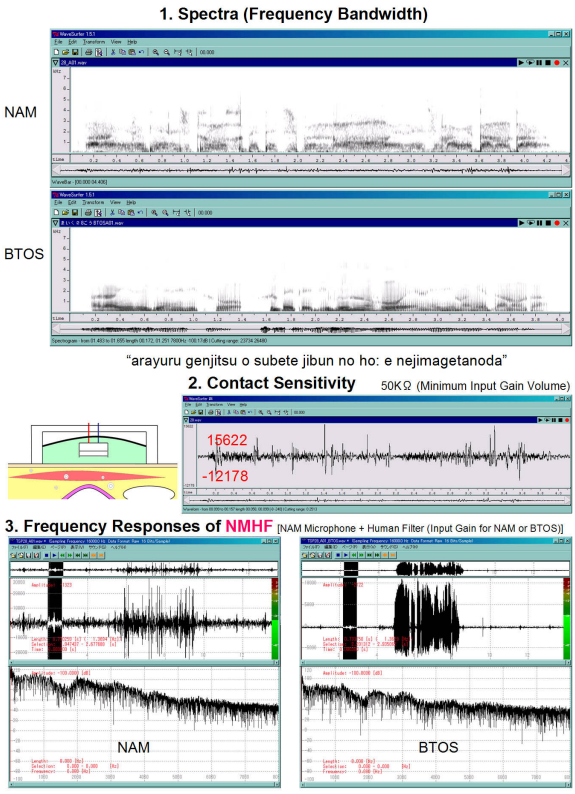


Figure 14: One of OCWSS types

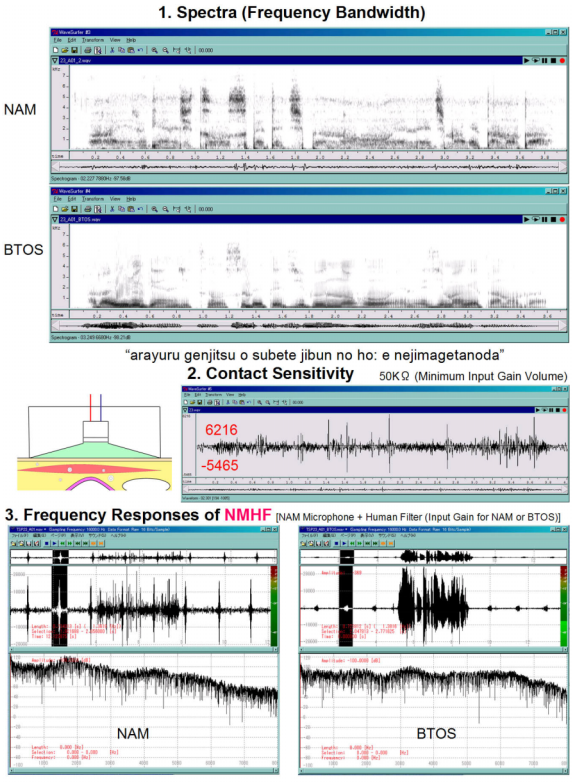


Figure 13: One of OCMSS types

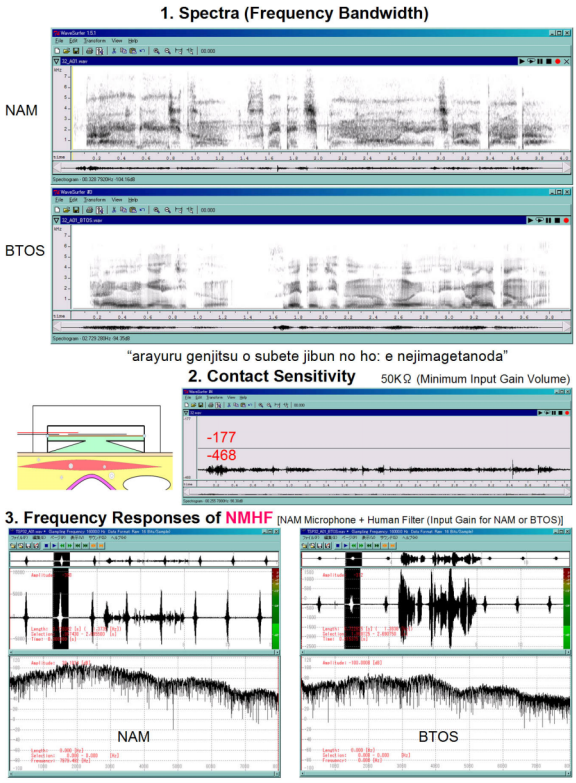


Figure 15: One of TMSS types

我々は数種類のシリコーンなどの素材とコンデンサマイクロフォンやセラミック圧電素子を用いて，多くの NAM マイクロフォン試作品を作成したが，その特性や性質から NAM マイクロフォンを新，旧合わせて大きく 5 種類に分類し，図 16 に様々な NAM マイクロフォンタイプの分類別の大まかな特徴を参考のため表にまとめてみた．

	ST	OCWHS	OCMSS	OCWSS	TWSS
Band Width	0-2KHz	0-4KHz	0-6KHz	0-4KHz	0-7KHz
Contact Sensitivity	middle	middle	middle	high	low
Noise Robustness	low	middle	middle	high	low

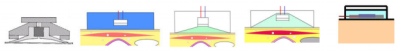


Figure 16: Comparison of NAM Microphones

5. 新 NAM マイクロフォンによる認識率

ソフトシリコーンを用いた新 NAM マイクロフォン三種のうち，OCMSS や TMSS に比し，接触面感度や外部雑音耐性には優れるが，帯域の最も狭い OCWSS タイプの NAM マイクロフォンを用いて，NAM 発話による大語彙連続認識実験を行った．特定男性話者の NAM 発話による新聞記事読み上げを 16KHz サンプリングし，通常音声男性不特定話者の Phonetic Tied Mixture (PTM) モデル[5]に HTK[8]を用い話者適応を行った（Iterative MLLR）．認識エンジンは Julius[5] を用いた．認識率の評価は Japanese Dictation Toolkit[5] を用い，評価用の 24 文を NAM 発話にて同じマイクロフォンにて読み上げて認識率を計算した. 図 17 にその結果を聴診器型 NAM マイクロフォンと比較して提示する．認識精度は聴診器型に比し約 5%上昇した．

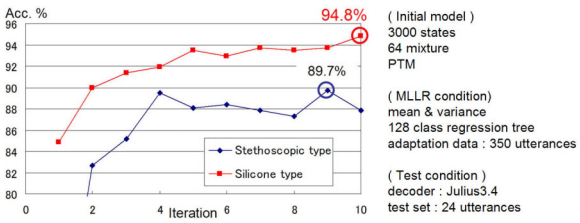


Figure 17: Iterative MLLR

6. まとめ

ソフトシリコーンを音媒体とする新しいタイプの NAM マイクロフォンは，上限 2ＫＨｚだった聴診器型に比し，より広い帯域を持ち，認識率を大きく向上させた．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った．また可塑性を持つ素材を選択したことにより，マイクロフォンの形状・デザインと大きさが，聴診器型に比べてより自由に設計できるようになり，小型化・薄型化が可能になっている．

現時点でまだ統計的な聴覚実験は行っていないが，携帯電話の帯域である 4KHｚ以上に帯域が広範化したことにより，ソフトシリコーン型ＮＡＭマイクロフォンでサンプリングされた NAM 音は，信号処理をせずに単に増幅して聴取しても聞き取り可能となった．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何らかの信号処理を加えて通常音声化したり，無声音のままでもより聞き取りを容易にしたものを直接通信に用いることが可能と考える．周囲の状況に気兼ねしない，いわゆる「無音声電話」などへの応用が期待できる．

現時点での課題としては，帯域，接触面感度，外部雑音耐性について，すべてに優れる決定版とも言うべき NAM マイクロフォンを作成することである．それにはコンデンサマイクロフォンの種類，セラミック型圧電素子，ピエゾ素子の種類とその固定方法を検討すること，また皮膚からの音媒体としての素材の種類と，硬さや弾性を変えての実験も必須である．加えて汎用インターフェースとして実験室や日常での使用の簡便さと，デザインやアクセサリーとしての装着時外観も考慮して，NAM マイクロフォンの皮膚への固定方法を考えていく必要がある．現在，携帯電話や PC，カーナビゲーションシステム等との規格化されたデバイス間無線通信も検討中である．これらのことをふまえ，今後 NAM を用いた研究を広く一般に可能にするため，日常的に室内や屋外で，また実験室で誰にでも気軽に使える標準化された NAM マイクロフォンデバイスを作成し量産す

ることが目標である．

また認識においても，現在の認識率の評価は，通常音声の不特定話者の音響モデル（PTM モデル）を NAM 発話サンプル文で話者適応（Iterative MLLR）して NAM 音響モデルを作成している．これを純粋に NAM 発話文のみからなる純粋 NAM 音響モデルの不特定話者 PTM モデルを作成することも，NAM に特徴的なパラメータ抽出法の考案と会わせて今後の課題である．

携帯電話というコンピュータが万人に普及し，ユビキタスコンピューティングが夢物語ではなくなり，また流れとしてウェアラブル・コンピューティングがこの先に見えつつある今ほど，その入力デバイスのインターフェースとしての質が問われる時代もない．今までコミュニケーションの道具として人間が使ったことのなかった NAM を幅広く，人対機械，人対人の新たなインターフェースとして用いることを提案し，ハンズフリーの音声認識や音声信号処理の豊かな技術蓄積を生かしつつ，しかも周囲に気兼ねしない，また環境の制約を受けにくい NAM というインターフェースを用いたコミュニケーションを NAM Interface Communication と名付ける．

参考文献

[1] 中島淑貴，柏岡秀紀，鹿野清宏，ニックキャンベル(2003)「微弱体内伝導音抽出による無音声認識」『日本音響学会 2003 年春季研究発表会講演論文集』pp.175-176.

[2] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell, “Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin”, Proc. ICASSP, 2003.

[3] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell, “Non-Audible Murmur Recognition”, Proc. EUROSPEECH, 2003.

[4] 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, 2003.

[5] 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. 56, pp. 255-259, 2000.

[6] C. J. Leggetter, C. Woodland, “Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models”, Computer Speech and Language, Vol. 9, pp. 171-185, 1995.

[7] P.C Woodland, D. Pye, M.J.F. Gales, “Iterative Unsupervised Adaptation Using Maximum Likelihood Linear Regression”, Proceedings of ICSLP, pp. 1133-1136, 1996.

[8] S.Yong, J.Jansen, J. Odell D. Ollason, V.Valtchev and Phil Woodland, The HTK Book, 2000.

[9] 竹原靖明(1991)『腹部エコーの ABC』（日本医師会生涯教育シリーズ）東京：医学書院.

[10] Helmut Ferner 編(1966)『臨床応用局所解剖図譜』(第１巻　頭部・頸部）東京：医学書院.