NAIST-IS-DD0361026

博士論文

NAM インターフェース・コミュニケーション

－その基礎としての肉伝導音声センサー開発と検討－

中島 淑貴

2005 年 2 月 2 日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士（工学）授与の要件として提出した博士論文である.

論文番号：　NAIST-IS-DD0361026

提出者：　中島 淑貴

審査委員：　鹿野 清宏 教授
横矢 直和 教授
Nick Campbell 教授
柏岡 秀紀 助教授

NAM インターフェース・コミュニケーション*

－その基礎としての肉伝導音声センサー開発とその検討－

中島 淑貴

内容梗概

非可聴つぶやき（Non-Audible Murmur: NAM）は「気導音としては周りが聞き取れないほどの無声音のつぶやき」の「肉伝導音」であり，音響学的には「声帯振動ではなく気道の乱流雑音を音源とする無声呼気音が, 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．音声の生成系である人体表面から直接 NAM をサンプリングすることにより, 高感度で聴取可能な音声信号として捉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．

第一に聴診器接着型 NAM マイクロフォンを開発し, 肉伝導する NAM をサンプリングして認識するのに適した装着位置を見つけた．HMM 音響モデルに EM 学習や話者適応を行って NAM 音響モデルを作成し，大語彙連続認識実験を行い，いわゆる「無音声認識」（非可聴つぶやき認識）の実用可能性を見いだした．またこの NAM マイクロフォンによりサンプリングされる体内伝導通常音声(Body Transmitted Ordinary Speech: BTOS) による BTOS 認識についても検討した．

第二に NAM 音の信号処理による通常音声化，いわゆる「無音声電話」などへの応用が考えられるが，聴診器型 NAM マイクロフォンによる NAM は

* 奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文,
NAIST-IS-DD0361026, 2005 年 2 月 2 日.

2KHz 以上にフォルマントが見られない．このため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た新型 NAM マイクロフォンを発明，開発し，NAM 音の帯域の広範化とともに接触面感度や外部雑音耐性の上昇を得た．このソフトシリコーン型 NAM マイクロフォンにより，NAM や BTOS をサンプルに用いた HMM による機械認識においても，人間による聞き取り試験においても，聴診器型に比し，その認識率が向上した．また他社製の肉伝導音声収録用センサーとの比較も行った．

第三として NAM マイクロフォンを同側で縦に 2 つアレイ化して装着し，ピッチ変動に伴う喉頭の上下動をパワー比により移動音源定位することで，F0 とは異なった視点から BTOS や NAM 発話のピッチを推定できる可能性を論じた，また音声の研究において人体を肉伝導の音場と捉える考え方を紹介した．

この NAM とその汎用音声入力インターフェースとしての利用価値の発見により，NAM を肉伝導の第二の音声言語として，その信号に既存の音声信号処理技術の蓄積を応用すれば，周囲環境に気兼ねせず影響も受けにくい，人対機械，人対人の新しい発話入力インターフェース・コミュニケーションが可能となる．これを NAM インターフェース・コミュニケーションと名付けて提唱し，その技術の根底の基礎となる肉伝導音声センサーの開発とサンプリング方法について検討を行った．

キーワード

インターフェース，非可聴つぶやき（Non-Audible Murmur：NAM），肉伝導，無音声認識，無音声電話，NAM マイクロフォン，体内伝導通常音声（Body Transmitted Ordinary Speech: BTOS）

NAM Interface Communication*

-Development and evaluation of flesh conduction voice sensors as the basis-

Yoshitaka Nakajima

Abstract

Non-Audible Murmur (NAM) is a non-voiced speech sound, created by turbulent airflow generated in the glottis and articulated in the vocal tract by speech-like movements of the tongue, lips, and jaw. It is similar to whisper, but is generally inaudible to persons other than the speaker. It can be detected by use of a skin-mounted microphone worn below the ear.

By sampling NAM signals directly through the thin layer of flesh at the top of the neck, the speech-like sounds can be converted into audible speech. By using a stethoscopic microphone, external noises can be shielded, yielding a speech source that is robust in noise, inaudible to nearby listeners, and suitable for recognition using a suitable retrained but otherwise standard speech recogniser. This thesis presents motivation for the design and placement of the NAM microphone, and presents results of large-vocabulary speech recognition tests using NAM speech. NAM speech is compared with Body Transmitted Ordinary Speech (ordinary voiced speech transmited through the flesh) sampled with a NAM stethoscopic microphone, and recognition results are presented for each type of speech.

As a prototype application for such non-voiced speech, the thesis describes a 'silent-speech-phone', where the NAM signal is rendered audible by signal processing, yielding clear formant information up to 2KHz. Several different designs of microphone were tested, using soft silicone, which has an acoustic impedance close to

* Doctor's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0361026, February 2, 2005.

that of human flesh, as an acoustic-damping material. This resulted in greater wideband sensitivity and higher contact sensitivity which served top increase robustness against external noise, and significantly improved recognition accuracy.

In order to increase sensitivity to prosodic information for this speech sensing technique, as developed and tested a further design, using the stereo signal generated by a pair of vertically-mounted NAM microphones. This enabled us to measure changes related to fundamental frequency that arise from movements of the larynx.

We propose NAM speech as a new all-purpose voice input interface and present speech signal-processing algorithms that allow this speech source to be used both for human-to-human and human-to-machine communication which is robust to noisy environments yet unobtrusive even in a quiet room where other people may be present.

Keywords:

Interface, Non-Audible Murmur (NAM), flesh conduction, NAM recognition, Non-Voice Phone, NAM microphone, Body Transmitted Ordinary Speech (BTOS)

目次

図目次

表目次

略語・新語リスト

当論文には新造語が極めて多く登場するので，参照に便利なように，ここにまとめた．*記号の付いたものは新語である．

*NAM (Non-Audible Murmur)：非可聴つぶやき

気道の乱流雑音を音源とする無声呼気音が，発話器官の音響的フィルタ特性により調音されて，肉伝導したもの．NAM 発話は行為としては「微弱なささやき声」と同じだが，あくまで肉伝導音としての立場からみた言葉．

*NAM マイクロフォン

NAM を体表からセンシングする目的で設計された体伝導音センサー．大きく分類すると聴診器型，ハードシリコーン型，ソフトシリコーン型などがある．現在はソフトシリコーン型が主流．

*BTOS（Body Transmitted Normal Speech）：体内伝導通常音声

NAM マイクロフォンでサンプリングする通常音声．「ビートス」と読む．

*肉伝導

人間の皮膚，筋肉，結合組織，脂肪組織などの軟部組織，いわゆる「肉」を音の伝導媒体とすること．

*肉伝導音声

NAM や BTOS などの，肉伝導音をサンプリングして得られる人間の音声．

*聴診器型 NAM マイクロフォン

肉伝導音を聴取するための聴診器の原理を応用して考案された，皮膚とコンデンサマイクロフォンの音媒体は微小密閉反響空間の空気である．

*ハードシリコーン型 NAM マイクロフォン

皮膚とセンサーとの間の音媒体にプラスチック～硬い消しゴムくらいの硬さのシリコーンを用いた NAM マイクロフォン.

*ソフトシリコーン型 NAM マイクロフォン

皮膚とセンサーとの間の音媒体に人間の肉の柔らかさと弾性に近いソフトシリコーンを用いた NAM マイクロフォン. 設計の基本発想やセンサーの違いなどにより大きく OCMSS 型，OCWSS 型，TMSS 型の三種類に分類される．現行 NAM マイクロフォンの主流.

ECM (Electret Condenser Microphone)

通常のコンデンサマイクロフォン.

*OECM (Open Electret Condenser Microphone)

ECM の振動電極板を露出させたもの．現在の所，手作業でこれを行う.

*OCWHS 型 (Open Condenser Wrapped with Hard Silicone Type)

ハードシリコーン型 NAM マイクロフォンで，センサー部に OECM を用いてハードシリコーンで全体を包んだもの．雑音耐性に優れる.

*OCMSS 型 (Open Condenser Mediated with Soft Silicone Type)

ソフトシリコーン型 NAM マイクロフォンで，センサー部に OECM を用い振動電極板と皮膚との間をソフトシリコーンで媒介したもの. よく使われるので簡単に「M 型」と呼ぶこともある.

*OCWSS 型 (Open Condenser Wrapped with Hard Silicone Type)

ソフトシリコーン型 NAM マイクロフォンで，センサー部に OECM を用いてソフトシリコーンで全体を包んだもの. 接触面感度が抜群である．これもよく使われるので，簡単に「W 型」と呼ぶこともある.

*TMSS 型 (Transducer Mediated with Soft Silicone Type)

ソフトシリコーン型 NAM マイクロフォンの一種で, センサー部に圧電素子を用いたもの，帯域は広いが，接触面感度が低い.

TSP (Transient Signal Priority) 信号

インパルス応答を測定するための基準化された信号. 高い周波数から低い周波数まで linear に時間変化する. 気導マイクロフォン特性などを測定.

*NMHF (NAM Microphone with Human Filter)

NAM マイクロフォンをはじめとする体伝導音マイクロフォンを人間の頭部に装着した状態を，ひとつの大きな仮想気導マイクロフォンとみなす考え方．この気導音感度が低いほど，実用時の外部雑音に頑強であると言える.

*LEI (Laryngeal Elevation Index)曲線

超音波診断装置で見たピッチの上下に伴う喉頭の上下動を, 甲状軟骨下縁のラインの上下動として時系列表示したもの.

*SOL (Stereophonic Orientation of Larynx)法

NAM マイクロフォンを縦にアレイ化することにより，ピッチ変化に伴う喉頭の位置を相対定位する手法のこと.

*Up/Dp パワー比

縦アレイ NAM マイクロフォンにおいて, 上 NAM マイクロフォンのパワーを Up，下 NAM マイクロフォンのパワーを Dp としたときの比率．F0 とは異なった次元のピッチに関する情報が得られる可能性がある.

第1章 序論

「あなたは、祈る時には自分の奥まった部屋に入りなさい。そして隠れた所におられるあなたの父に祈りなさい」（マタイの福音書 6:6）

1.1 まえがき

声を出すことなく，したがって人に聞かれることもなく，自分の意図した相手にだけ（それが人間であっても機械であっても）リアルタイムに意志を伝達することができるとしたら，それはいわゆるテレパシーであろう．しかし思考内容がすべて伝わってしまうとしたら，社会生活では困ることの方が多い．口の中で小さくつぶやいた，周りに聞こえない声で「伝えたいこと」だけが伝って，意図した人とだけ会話ができたり，ロボットや車が動き出したりすれば，それは素敵な魔法である．そしてそれは現在の 21 世紀初頭のインフラや音声信号処理を中心とする科学技術に,「発想の転換」を加えれば実現可能であり，22 世紀を待つ必要はない．

白隠禅師の禅の公案のひとつに「隻手の音声（せきしゅのおんじょう)」というものがある．両手をポンと打つと音が出るが,「片手だけの音を聞いてこい」という有名な無理難題である．

音声信号処理の研究は，音の空気伝達を気導マイクロフォンで捉えて，そこから始まるものが多かった．人間は音声を耳で聞くから，気導音声の世界から研究が始まるのは当然である．しかし空気だけにとらわれすぎるならば，それは「隻手の音声」ではなかろうか．人間の音声は肉と空気の複雑な相互作用，二つものの関係性の中から生まれる．

鐘（かね）と撞木（しゅもく）のいったいどちらが鳴っているのかというのは馬鹿げた問いである．古い都々逸（どどいつ）にもこう唄われている．
「鐘が鳴るのか撞木が鳴るか鐘と撞木の合いが鳴る」

1.2　研究の背景

ユビキタス・コンピューティングやウェアラブル・コンピューティングが声高に叫ばれる中，入力インターフェースに何を使うかという問題は，目下の大きな課題である．ハンズフリーである音声入力は，その期待に応えるものとして注目されている．しかし「ユビキタス」が意味する「あらゆるところで」は，ユーザーが「物理的にあらゆるところ」でコンピューティングできるという意味だけではなく「周りの他者との関係性の中でのあらゆるところ」という意味でなければならないと思う．

音声を入力インターフェースとして使う技術として，人対人の遠隔コミュニケーションの道具として 120 年以上の歴史を持つ電話がある．しかし近年携帯電話の爆発的な普及で，電話の発明者達が予想もしなかったような場所や場面で，電話が使われるようになった．また人対機械のコミュニケーションの道具としては，約 30 年の技術蓄積をもつ音声認識技術がある．技術的には十分実用段階にあり，安価で市販アプリケーションの入手も可能であるが，何故か人々の日常生活には，全くといっていいほど普及してない．

電話が携帯電話として現に日常のあらゆる場所で使用されるようになり，音声認識のアプリケーションを，オフィスや屋外など公共の場で実際に使用していこうとすると，その外部雑音対策や，公衆の面前で声を出すことの弊害も，周囲環境や公共性とのバランスから，真剣に考慮する必要がある．

騒音環境下での雑音対策として，骨伝導を主とする体伝導音声の研究は国内外でいくつかあるが[52]，これらはすべて通常音声をその対象としたものであった．市販の骨伝導スピーカーを用いた携帯電話もこの範疇に入る．

また微小発声，無声音発声の研究も見られており[3][16][56][57]，入力シ

ステムとして用いようとした研究もある[59][60]．ただしこれらはすべて気導音収録を前提としている．しかしきわめて微小な無声音発声を収録しようとすれば，当然増幅率を大きく上げざるを得ず，同じ気導音である外部雑音に対して通常音声の場合よりさらに脆弱となる．

さらに全く無発声で，発話時の口周囲の動態を，顔面の筋肉の筋電図から読み取り，発話内容を認識しようとする試みもあるが[30][31][44][45][46]，現段階では五母音の識別の段階にとどまっている．

1.3　研究の目的

本研究の目的は，人間や機械に対しての新たな音声情報伝達手段として，「周りの人に聴き取れないような，声帯の振動を伴わないつぶやき声」を，気導音としてではなく，むしろ気導音を排除して「体内伝導音を高感度，広帯域で効率的に収録するために新たに開発した体表密着型センサー」により音声の生成系から直接肉伝導音としてサンプリングし,「既存の音声信号処理の技術」を応用することによって, 外部雑音に頑強で, 周囲に気兼ねしない, ユビキタスかつユニバーサル・デザインの新たな発話入力インターフェースを実現するための基礎を築こうとするものである．

具体的には，その技術を音声認識や電話に応用することにより，周りの人に聞こえない，いわゆる「無音声認識」や「無音声電話」などを実現することであり，声帯の振動が不可能な発声に障害を持つ方々を補助するための礎を築くことである．また理念としては，人を点としての音源とみなして，そこから放射された気導音声のデータ解析をすることが主流であった「隻手音声」の音声研究に，肉媒体の音場，肉伝導音声の豊かな広がりを示して一石を投じ，波紋を広げることである．

そのための一番の根底であり，このインターフェース開発を展開していくための基礎としての「肉伝導音声を高感度，広帯域でサンプリングできるセンサー」の開発と改良，そしてその評価をこの論文の主眼においている．

1.4　もうひとつの音声言語「NAM」

ではその「周りの人に聞こえないような，声帯の振動をともなわないつぶやき声」を何と呼んで，どう定義すればよいか．

人間の音声言語は，声帯を振動させて発生する音源が，調音器官の運動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，声門を著しく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[16]．どちらも距離の差異こそあれ，「空気を媒体とする他者への音声情報の伝達」を目的として発声する音声言語であり，これを第一の音声言語とする．

しかし我々の日常生活を思い起こして，自分たちがもうひとつの言語発話行動をしていることに気が付いて欲しい．人に聞かれないように口の中で独り言をつぶやくとき，また神社仏閣などで祈りや願い事をひそかに口の中で唱えるときの声である．それは周囲の人々ではない「ある何者か」に語りかけるための声であり，声帯を振動させない無声音であることにおいて「ささやき声」に似ているが，もっと微弱である．自分の願い事などは周りの人には聞かれたくないものであり，人に聞かせることを前提にしない，または人に聞かれたくない独り言であることにおいて「つぶやき声」に似ているが，無声音である．これがこの論文で問題とする第二の音声言語である．

辞書で探してみたが，この発話行動には日本語では適当な名前が付いていないので，これを「非可聴つぶやき（Non-Audible Murmur: 以下 NAM）」と呼ぶことにする．音響学的な定義は後述する．

「無声音つぶやき」や「独り言ささやき」などと呼んでもいいのであるが，「つぶやき」を「非可聴」にしようとすると無声音にせざるを得ず，また「聞こえない」ことに実用上の利点を認めて，力点を置きたかったからである．

この NAM 発話行動は，個人の内部で処理される音声言語活動であり，有史以来人間どうしのコミュニケーションの道具として使用されたことはなく，祈りの言葉の例に見るように，むしろ「人間以上の存在」にひそかに語りかける言葉に近かった．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．今まで NAM 発話は単にそのパワーの大きいものが,「ささやき声」としてごく近辺にいる人への限定されたコミュニケーションに使用されていたにすぎない．しかし気付かれていなかったのに行為としては存在したからこそ NAM 発話は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．

表 1.1 にさまざまな発話様式を，声帯の振動と情報の伝達意図という観点から，わかりやすいように分類を整理した．NAM 発話はつぶやき声とささやき声の「欠けた性質」どうしを合わせた物であると言うこともできる．

表 1.1　声帯の振動と伝達意図からみた音声言語の様々な発話様式

	通常音声	小声	つぶやき声	ささやき声	NAM
声帯の振動	○	○	○	×	×
伝達意図	○	○(限定)	×(独り言)	○(限定)	×(独り言)

1.5　NAM の定義

NAM 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」を指す造語として生まれたが，その収録方法や，後述する発見の経緯が，従来の気導音マイク収録によるものとは全く異なるため，その概念に「音の伝導媒体」も含める．

NAM の音響学的定義は**「声帯振動ではなく気道の乱流雑音を音源とする無声呼気音が，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．

つまり NAM とは「気導音としては周囲に非可聴な，調音無声呼気音の肉伝導」のことである．正確を期すため，音や信号であることを強調したいときには「NAM 音（NAM sound)」，「NAM 信号（NAM signal)」，発話行動に力点を置きたい場合は「NAM 発話」や「NAM 発声」などと表記することにする．「音声」や「ささやき声」などの言葉は「気導音」を前提としており，発話行動自体を指す場合もあれば，音を示すこともあり，また収録されたデータそのものを指すことも文脈によってはあるので，NAM という言葉もそれと同様である．「気導 NAM 発話音」などという表現もありうる．

また「非可聴」とは言っても，どこまでが「非可聴」なのか線引きが難しく，また距離や周囲雑音環境によっても聞こえないレベルは大きく異なる．そのためここでは物理的な線引きをせず,「発話者本人が周囲の状況に応じて,周りに聞かれたくない意図から無声音で発話したもの」を NAM 発話とする．騒音環境下では「大きなささやき声」の肉伝導も NAM になりうる．

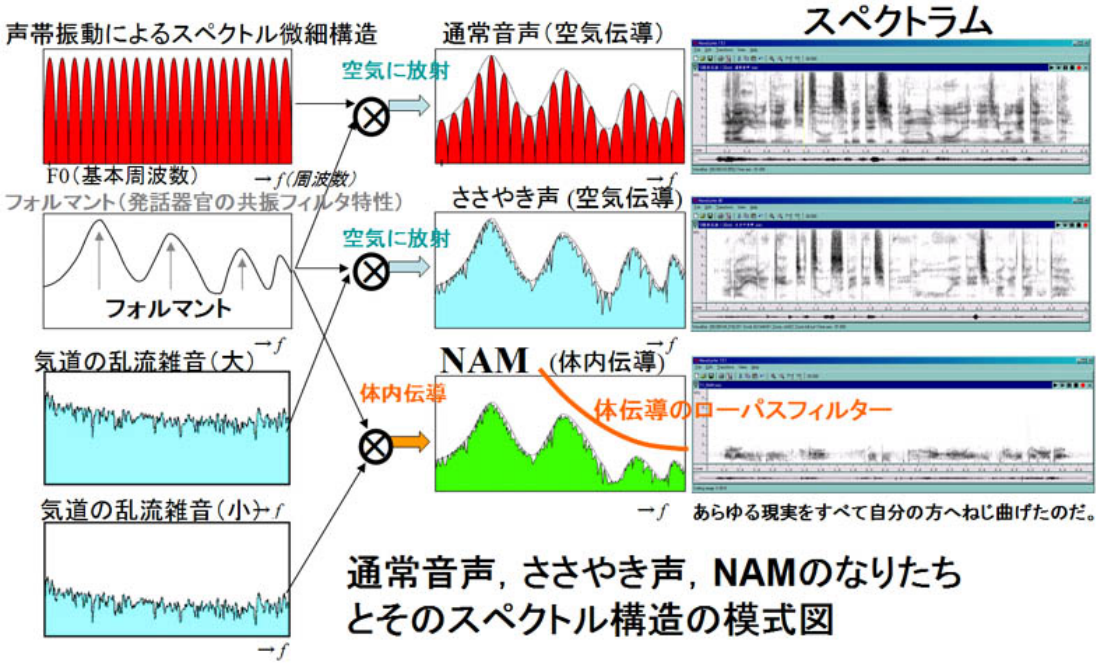


図 1.1　通常音声，ささやき声，NAM のなりたち

図 1.1 に通常音声，ささやき声，NAM の成り立ちの原理を模式的に表現してみた．人間や音声認識が聞き分けているのは，図で言うフォルマントの情報であり，そのフォルマントを形作るための素材である音源が，声帯の振動によるスペクトル微細構造であるか，気道の乱流雑音であるかの違いはあっても，音素によるフォルマント構造の違いはほぼ相似形である．通常音声を聞き分ける能力でささやき声も聞き分けることができるのは，そのためである．NAM の場合はこれに肉伝導によるローパスフィルターがかかって高域のフォルマントが消失もしくは変形するが, やはりある種の声に聞こえる．

たとえて言うならば，大きな石でできた精密な仏像も，小さな木の粗彫りの仏像もどちらも「仏さん」として人々に認識されることと同じである．

1.6　NAM マイクロフォンとは何か

「NAM を体表からセンシングする目的で設計された体伝導音センサー」を「NAM マイクロフォン」と定義する．これを図 1.2 のごとく耳介後下方部の皮膚に密着させて NAM 音を拾う．この位置はほぼ口と同じ高さであり，音響管を骨の遮蔽なしに肉だけを介して後方から覗く位置にあたる．音伝搬の媒体は，主に頭部の軟部組織，いわゆる肉伝導である．

次章以降で詳述するが, NAM マイクロフォンは聴診器型に始まって, 様々な形態がある．現行開発モデルとして，コンデンサマイクロフォンの振動電極板と皮膚との間に音媒体として人間の筋肉や皮膚とほぼ同じ音響インピーダンスを有するソフトシリコーンを用いる，ソフトシリコーン伝導型などがある．これによって空気の伝搬を介した音声ではなく，音声生成系から直接振動を振動電極板に伝える効果がある．肉伝導により音声をサンプリングすれば，同じ増幅率のマイクアンプを使っても気導音よりはるかに大きいパワーの信号を得ることができる．NAM と通常音声ではその気導音としてのパワーは数百〜数千倍異なるが，NAM マイクロフォンで NAM をサンプリングすれば，気導音としての通常音声を通常マイクロフォンで収録する場合に

比して，6〜10dB 感度が高まり，むしろマイクアンプの増幅率や出力レベルを低下させても聴取に十分な信号をサンプリングできる（第三章 3.11 参照）．

図 1.2　NAM マイクロフォンの一例（CEATECH 2004）

骨伝導マイクロフォンや Throat マイクと呼ばれるセンサーなども，体伝導音をサンプリングするという点で NAM マイクロフォンに似ているが，あくまでその本来の目的は，「通常音声の収録」であり，NAM をサンプリングできる感度には設計されておらず，かなり増幅率の高いアンプを使わないと NAM は信号として現れない．一般に増幅率を上げれば上げるほどその音質は悪くなり，外部雑音に対しても脆弱となる．

NAM マイクロフォンを使用すると，マイクアンプの増幅率や出力レベルを適切に低く設定すれば，もちろん肉伝導の通常音声も収録可能である．「NAM マイクロフォンによりサンプリングされる通常音声」を「体内伝導通常音声(Body Transmitted Ordinary Speech: 以下 BTOS) 」と定義する．

NAM と BTOS など，NAM マイクロフォンでサンプリングされる音声を「肉伝導音声」と呼ぶことにする．肉伝導音声を NAM マイクロフォンのような接触型マイクロフォンで収録することの利点と欠点であるが，利点とし

ては，まず前述のように感度の面から，通常音声を気導通常マイクロフォンで収録するよりはるかに大きなエネルギーで音声をサンプリングできることである．自作の NAM マイクロフォンでは，通常音声を収録する増幅率かそれ以下で NAM を収録可能であり，つまり BTOS を収録する場合はさらにマイクアンプの増幅率を絞ることができる（またはマイクアンプ不要か，増幅率 0dB でも収録可能）．つまり目的音である肉伝導音声は小さな増幅率で大きく収録できるということである．しかも不必要な外部雑音は体のフィルタを通すこととアンプの増幅率を下げることの相乗効果で低減させることができるということ．この二点より，外部雑音の混入を避けて，NAM のように微小な音声をサンプリングするのに優れた音声収録方法であると言える．

欠点としては体表からサンプリングするため体のローパスフィルターの特性が働くことと，音響管の終末端である口唇の放射特性がほとんど入らないということから，NAM 信号や BTOS 信号の帯域が気導音声と比して狭くなるということである．つまり，全体的に「こもった音」に聞こえる．

1.7　ささやき声と NAM

ささやき声と NAM とはどう違うのかという質問がよくなされる．確かに実際になされる行為としては，NAM 発声と「微弱なささやき」はほぼ同義である．つまり「気導 NAM 発声音」が「微弱なささやき声」である．

NAM 発声とは本来，声帯を振動させたり，気道を狭めたりすることもなく，ほぼ呼気に伴って口だけ動かすようなものを言うが，声道の狭めが強く呼気量が多ければ多いほど乱流雑音のパワーが上がり，それは「ささやき声」に近くなる．しかし聞かせたくないという意図が強ければ，自然に呼気量と気道の狭めは小さくなる．行為としては，そのパワーによって気導 NAM 発声音を soft whisper と呼んでも「微弱なささやき声」と呼んでもかまわない．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」と気導 NAM 発声音との間に，パワーの大小という大まかな違いはあ

るものの「行為として」は正確な物理的境界線を引くことは実際問題として難しい．通常音声の発話時と同じような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，声門の狭めによる強い乱流雑音を音源としているものまで，NAM 発声にもかなりのバリエーションがある．このバリエーションは周囲環境，特に環境雑音や周囲にいる人との距離に依存することが多い．実際，「行為としての」NAM 発声は「聞き取れないほど微弱なささやき声」であり，実験の際など，人に初めて NAM 発話をしてもらうときに，NAM 発声という行為の理解を容易にするため，「聞こえないぐらい静かにささやいて下さい」と説明することもある．

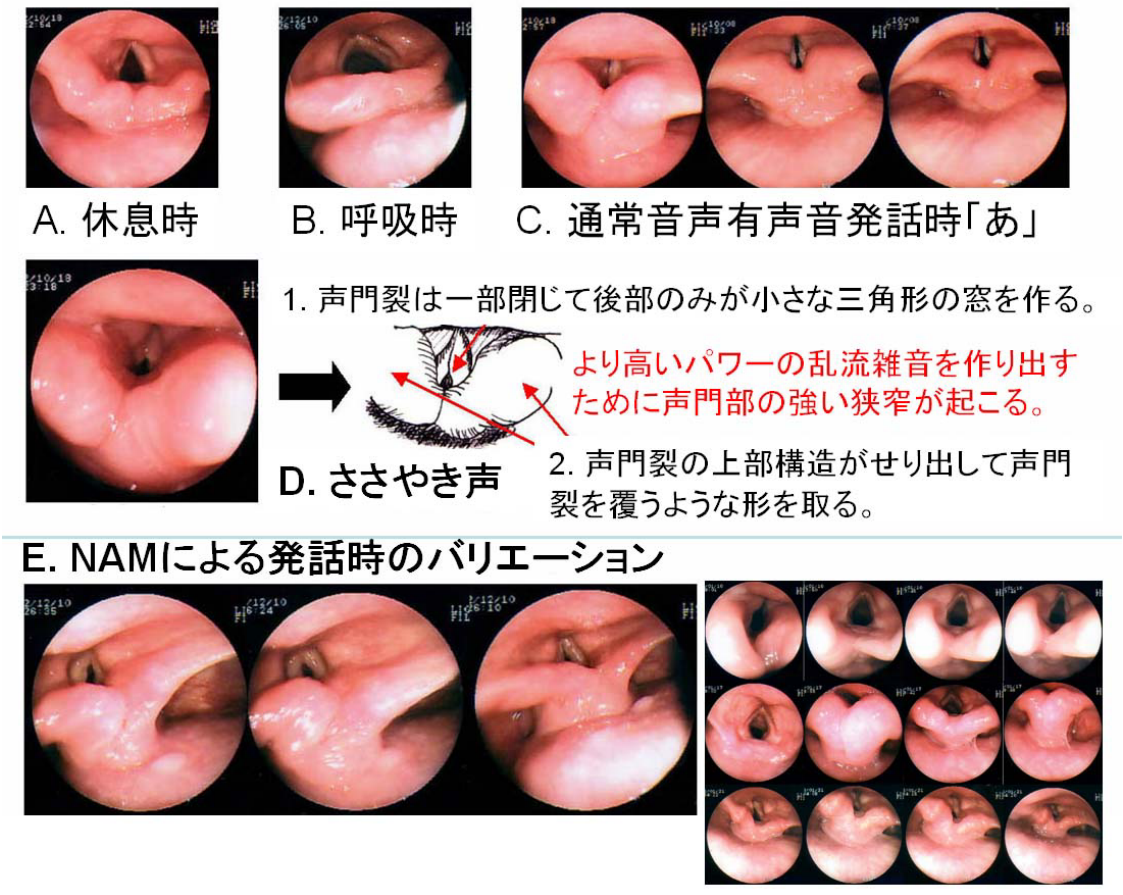


図 1.3　様々な発話時における声門部の内視鏡像

図 1.3 に 5 名のボランティアの様々な発話時における声門部の内視鏡像を示す. D が一般的なささやき声を発するときの典型的な声門部の形態であり. 文献にも記載されている通り，二通りの狭窄が声門部に起こっている．声門裂が一部閉じて小さな三角形の窓を作ることと，声門列の上部の肉が上にせり出して声門裂に覆い被さるような形をとることである．E に NAM 発話時の声門裂の形態を集めてみたが，ささやき声発話時のこの典型的形態をとったものは観察した中では一例もなかった．声門裂の狭めはあるが，上部構造による狭めはないもの（右の三つの写真），ほとんど呼吸時と変わりないもの（左上段）上部構造の盛り上がりは観察できるが，声門裂は開いたままのもの（左中段），どちらも中途半端に起こっているもの（左下段）など，個人差とその時の NAM 音量によるバリエーションが大きい．言えることは NAM 発声時には典型的なささやき声に比べて，声門部の狭めが中途半端である．

音源が声門部の狭めによる乱流雑音であるということにおいて，NAM 発声とささやき声との物理学的，解剖・生理学的な面での境界は曖昧であり，行為としては共通部分が多いが，NAM という概念と「ささやき声」との大きな差異は，その発話者の「伝達意図」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「ささやきたい」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 発話は人に聞こえないように，または聴者を想定せず，内部処理される発話行動である．

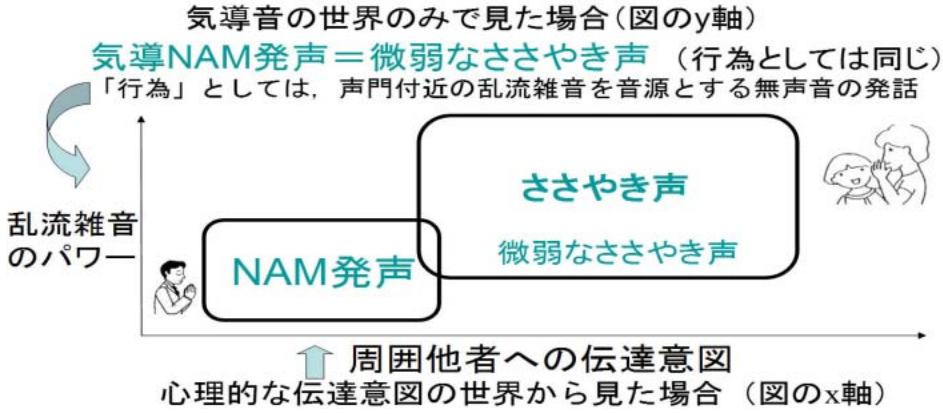


図 1.4　行為としての NAM 発声とささやき声発声

NAM と「微弱なささやき声」との決定的な差は，その「伝導媒体」である．「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[3][32][56]．この点で人間の軟部組織，つまり肉を伝導したものであると定義した NAM とは本質的に異なる．肉と空気の複雑な相互作用で無声音声は発生するが，それを空気の側の世界から捉えたものが「微小なささやき声」であり，肉の側の世界から捉えたものが NAM であると言える．

言い方を変えれば，気導音の世界ではコミュニケーションには使えそうもなく顧みられかなった微弱なささやき声が，肉伝導音の世界では極めて高感度に捉えられることがわかり，初めて「使用に足る声」であると認識されたのであって，対象となる行為は同じであっても，それを見る視点の違いは大きい．NAM の定義に「伝導媒体」を含めてあるのはそのためである．図 1.5 にその概念図を示す．

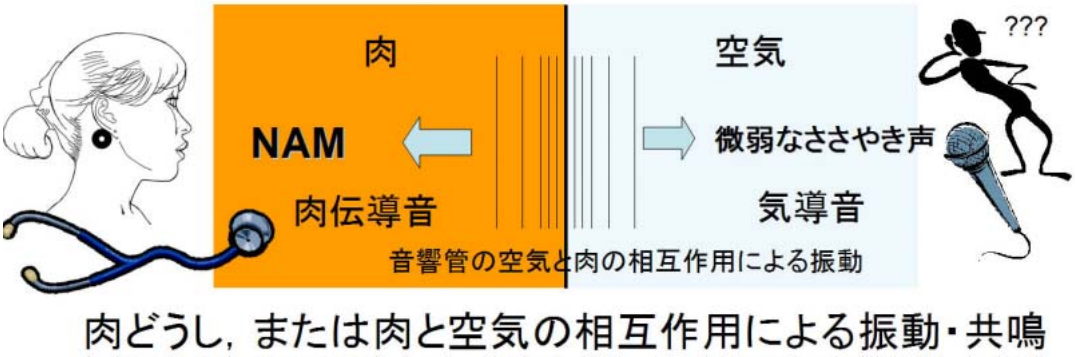


図 1.5　NAM と「微弱なささやき声」の伝導媒体の違い

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，これは呼気のまったく伴わない NAM 発話であるとも言える．これが可能であれば本当の意味での無音声認識である．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い．無論長い文章は発話できないし，発話，無発話のオン・オフも判別が困難である．

1.8　骨伝導と肉伝導

骨伝導とはどう違うのかという質問もよくなされる.「骨伝導」という言葉は元来, 聴覚障害者向けに開発された「骨伝導スピーカー」に由来している.

クジラは下顎の骨で，水中を伝わる音の振動を内耳に伝えているし，ベートーベンは指揮のタクトを口にくわえてピアノに当て、音の振動を歯から頭蓋骨を経て、内耳まで伝えることでピアノの音を聴いたと言われている。

人間が普段聴いている音には二種類あって, それは気導音と骨導音である. 気導音の場合，空気の振動が耳たぶで集められて耳の穴（外耳道）に入り、鼓膜を振動させる。この振動が中耳で増幅され、内耳のうずまき管内部のリンパ液中に浮かぶ聴覚神経の先端部が揺れ動くことで、人間は振動を音として認識している。それに対して骨導音の場合は、外耳や中耳を経由することなく、頭蓋骨内部に埋め込まれた内耳のうずまき管に直接音の振動を伝え、リンパ液中に浮かぶ聴覚神経が揺れることで、音が聴こえる。自分の耳を塞いでも、自分が発した声が聞こえるのは、この骨導音があるからである。

いずれにせよ気導音や骨導音を知覚するのは「内耳のうずまき管」であって，これが「骨の中に埋め込まれていること」が「骨伝導」の由来であり,「スピーカーという出力装置であること」と,「知覚である」ということにおいてその「骨伝導」という言葉は意味をもつ．肉に押し当てて振動を体に伝えていても, 最終的に骨を振動させなければ, 音を知覚できないからである. しかし「スピーカー」の逆で体から振動をセンシングする「マイクロフォン」を考えたとき，その「骨伝導」という言葉に意味はなくなる．世にある「骨伝導マイク」と呼ばれるものは,「肉」に押し当てて振動をサンプリングしているのであり,「肉伝導マイク」や「体伝導音マイク」と総称はできても「骨伝導マイク」と呼ぶ意味はあまりない.「骨伝導スピーカー」があまりにも有名であるため，その裏返し表現としての名称であると考えられる.

また世に言う「骨伝導マイク」は通常音声を雑音下で収録するために設計されたものである．内部構造のわからないものもあったが，ほぼ全部がセラ

ミック圧電素子かピエゾ素子を用いている．当然ながら NAM をサンプリングするために設計されたものはないため，通常音声よりはるかに低いパワーの NAM を収録できる感度のものは見当たらなかった．仮に増幅率を上げても，音質は自作 NAM マイクロフォンに及ばなかった（三章 3.10 参照）．

NAM マイクロフォンはその本来の発想が医療用聴診器である．日常的に聴診器を業務で使用していると，当たり前のように身に付くことであるが，経験的に聴診器はすぐ下に骨のある硬い部分にはあまり当てない．何故なら肉の中で起こっているイベントとしての音は，肉の軟らかい部分に当てた方がよく聞こえるからである. この理由として, 音は骨にも肉にも伝導するが，肉と骨の音響インピーダンスがあまりにも違うため，その両者の界面で音が反射減衰を起こすためである．

人間の音声を作り出す音響管は，歯を除いてそのほとんどが肉でできている．子音はそのほとんどが肉と肉のぶつかる音であったり，空気が肉と摩擦を起こす音であったり，その振動源は肉と空気の両方である．また母音は声道という音響管の共鳴であって，共鳴しているのは中に存在する空気と，音響管という肉の管である．骨も一部で音響管を形作ってはいるが，あくまで空気と接する部分は肉である．

いわゆる「骨伝導マイク」というのは，通常音声収録用途で，肉の下にすぐ骨のあるような場所（側頭部や耳孔など）に装着することを前提に設計された本来「体内伝導通常音声サンプリング用マイクロフォン」と呼ぶべきものである．そして元来「骨伝導」という言葉は「スピーカー出力」と「知覚（聴覚）」を概念の主体とする言葉である．

一方「NAM マイクロフォン」とは，音響管から骨の遮蔽なしに肉を通して収録できる場所に装着して，NAM をサンプリングできる感度に設計された「肉伝導 NAM サンプリング用マイクロフォン」である.「骨伝導」とは逆の「肉伝導マイク入力」と「微小音声発話」を概念の主体とする言葉である．

ただ骨伝導マイクと同様, 体内伝導通常音声である BTOS も収録可能なので,「肉伝導 NAM＋BTOS サンプリング用マイクロフォン」でもある．

1.9　NAM Interface Communication とは

この論文のタイトルでもある NAM Interface Communication とは何か定義しておく．NAM マイクロフォンで体表からサンプリングすることにより得られる NAM を中心とする体内音信号（BTOS や体伝導雑音を含めて）をインターフェースとして，それに現在発展を遂げつつある音声信号処理技術を生かすことで可能となる，人間の人間に対するコミュニケーション，人間の機械に対するコミュニケーションのことである．

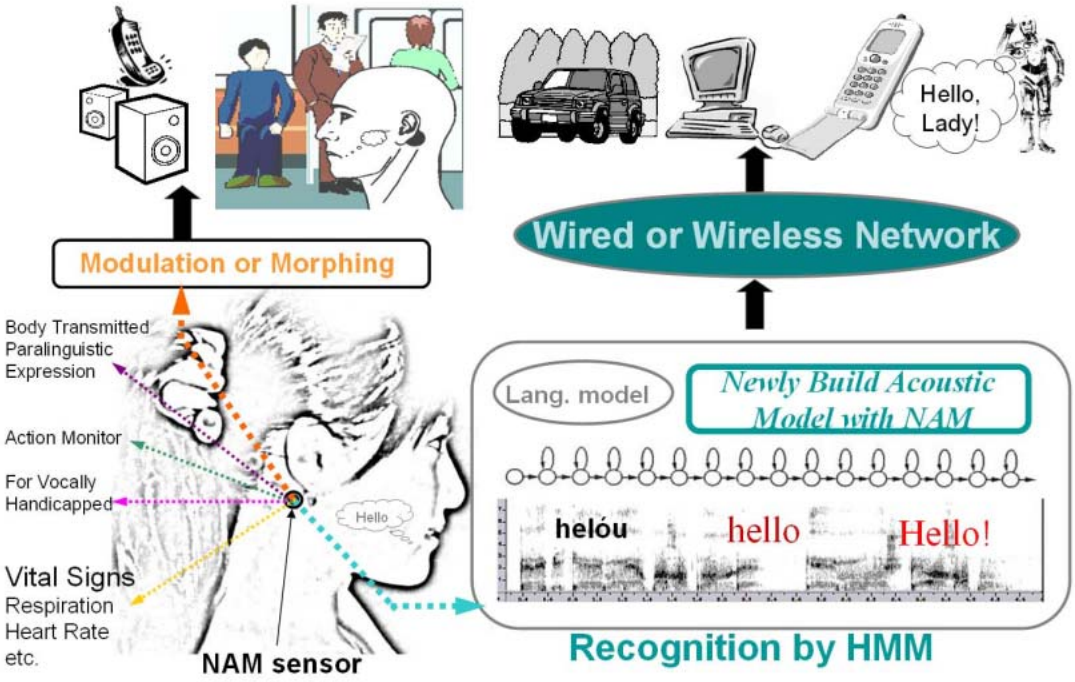


図 1.6　NAM Interface Communication の概念図

携帯電話というコンピューターが万人に普及した今，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，また周囲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（NAM Interface Communication）を提唱する．図 1.6 に簡単な概念図を示す．

実現すれば，音声を用いた入力の普及を加速させ，使用が周囲環境に束縛されない実用的で静かな音声入力インターフェースとなる．それでこそ「ユビキタス」という言葉が本当に意味を持ち，キーボードやテンキーよりはるかに多くの人々が使いこなせる「ユニバーサルデザイン」となる．

現在の所大きな流れとして二つの研究課題がある．一つは図右半分に描かれたいわゆる「無音声認識」で，NAM を既存の音声認識の技術を用いて認識しテキスト化する NAM 認識のプロジェクトである．もう一つは図左半分の上方に描かれたいわゆる「無音声電話」で，無声音である NAM を声質変換や音源付与など既存の音声合成の技術を用いて，通常音声化するプロジェクトである．この二つの大きな流れは NAM 関連分野の大きな二大潮流であることは間違いない．その他にも図の左端に小さく列挙したテーマは将来的に大きなテーマとなりうる項目である．

喉頭ガンなどで喉頭除去の手術後や神経筋疾患などの発声障害など，いわゆる声を失った人々に対する NAM 関連技術の応用は，健常者への応用以上に大切なテーマであり，これを実証してこそのユニバーサルデザインである．

またこの技術が普及して，NAM マイクロフォンを誰もが気軽にいつでも装着するようになったとき，感度も帯域にも優れた電子聴診器を 24 時間着けているのと同じ状態が発生する．NAM マイクロフォンには音声以外にも脈音や呼吸音など生体にとって重要な情報が常に入り，窒息や不整脈，心停止などは言うに及ばず，様々な生体音のモニタリング機能を持たせることができる．加えて様々な行動による特定の雑音パターンも行動モニタとして活用できる可能性があり，バイオメトリクス方面への応用範囲は広いと考える．

昔ならば，機械は人間とは離れた場所にある筐体のモノとして存在した．しかし現在のテクノロジーはそれをどんどん小型化し，機械は「身に着けられるモノ」となってきた．携帯電話はその使用方法が，昔の電話の受話器に似ていることから，まだまだ「電話」というイメージが抜けないが，CPU を備え，音声信号処理を内部で行う立派なコンピューターである．21 世紀初頭現在，個人が一人一台の情報端末であるコンピューターを身に着けて歩いているといっても過言ではない．送受話の部分に NAM マイクロフォンを使用し，小さく無線デバイス化すれば，携帯電話をそのコミュニケーションの主たる処理端末として使用することが可能である．

1.10 この論文の構成

2 章，3 章，4 章の内容は，筆者が奈良先端大在学中に行った大きな三つの仕事であり，出願した特許の三つの内容の骨子でもある．

- NAM の存在に気づき，それを体表から肉伝導でセンシングするためのデバイスを開発し，最適センシング位置を決め，音響モデルを作成して，NAM の大語彙連続認識が可能であることを示したこと（2 章）．
- 認識精度の上昇，無音声電話の実現に向けて高感度かつ高帯域で肉伝導音をサンプリングするためにソフトシリコーンを音媒体とした全く新しい体伝導音センサーを開発したこと（3 章）．
- NAM マイクロフォンを縦アレイ化することにより，喉頭の上下動を感知することによるピッチ予測の可能性を示し，音声を発する点として人間を捉えるのではなく，肉媒体の音場と捉える観点を示したこと（4 章）．

この論文は学術論文の通例の形式からははずれていて，それは自覚している．この世界の識者からは，「これは（自分たちが見慣れた）論文ではない」とお叱りを受けるかも知れないが，奈良先端大で自分の仕事としてやったことを，順を追って正直にまとめた．未来の後輩達が「こんなおもしろい博士論文がある」と目を輝かせてくれるよう，彼らに向けて書いたつもりである．

第2章 非可聴つぶやき認識の必要性とNAMの発見

2.1 はじめに

地元の図書館の閲覧室に,「音声で本が検索できます」と大書された書籍検索システム用の PC が数台並んでいる．あらゆる世代の利用者が，頻繁に検索を繰り返す様子を一時間ほど観察してみたところ，その装置に語りかけられることは一度もなかったし，試用してみる人も皆無であった．利用者の全員が,画面に触れるタッチパネル式のインターフェースを使用したのである．子供に頼んでこの音声認識システムを使ってもらったが，図書館という静音環境も手伝って，ほぼ確実に動作する．しかし自分で受話器を取ってみて初めて利用者の心理が理解できた．図書館では声が出しにくい．確かに閲覧室は，音声認識システムの実用の場に最適の低雑音環境である．音声認識入力もほぼ誤動作はない．しかし静音環境であるからこそ，声を出して周囲に迷惑な雑音を作り出すことに躊躇してしまう．何より書名やキーワードを発声することは,自分のプライベートな興味の対象を周囲に宣言することになる．一年後の現在，音声認識入力システムは撤去されている．

また音声対話による大規模知識ベース検索システムとして大学図書館内に置かれた音声認識入力可能なシステムでも，実際に入力されている利用者の音声の記録を調べてみるとささやき声が多かったという報告もある[29]．例えばパソコンのオンラインヘルプなども音声で利用できるようだが，自分が使う立場に立ったとしたら「ワープロはどうやって立ち上げたらいいです

か？」というような初歩的な質問は，周りに人がいる場では大声では聞きにくいということが心理的に理解可能である．

音声認識システムは，人間の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 年に渡ってその実現が試みられてきた．現在では隠れマルコフモデル（HMM)を用いた大語彙連続音声認識（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．認識精度においてコマンド認識は言うに及ばずディクテーションにおいても，室内静音環境ではもはや十分実用レベルにあるとさえ言える．しかしこの便利なフリーハンドの入力インターフェースを，日々実用している人を周囲に全く見かけないのは何故であろう．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．また音声認識システムの開発者自身が, 日常の入力に音声認識を使っていないことが多いのは何故なのか.

上述の図書館での実例は，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，周囲の人々に聞こえる声を明瞭に出さねばならないならば，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ことを示唆してはいないだろうか.

それは音声認識システムの機能面での不足と言うよりも，音声認識システムが内包する「実用上の本質的な欠点」が現在まであまり考えられたことがなかったためと考えられる.「音声認識の最大の欠点は, 声を出すことである」というのは逆説的で大胆な表現であるが，その本質を端的に表している．

音声認識はその大前提として，外部マイクロフォンを使って空気中に放散された気導音声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などでの使用を考えた場合，大きな欠点である．またオフィスや公共の場での使用を考えた場合，逆に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としたら，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．市販の音声認識アプリケーションを購入して使い始める際に，話者適応用の文章を数十文読み上げるのも，人がいるオフィスや研究室では不可能である．

加えて，音声認識を使ってみれば実感としてわかるが，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」ものである．特にそれが内容を匿秘したいものであれば尚更である．

携帯電話を使って電車内などの公共の場で会話しているのを見ても，我々はそれほど不自然に感じない．なぜなら普段から受話器を耳と口に当てて会話しているのを見慣れていて，対話している対象が想定できるからである．だがあらゆる機械に音声認識入力が普及したとしたら，受話器や携帯電話もなく空中に向かって声を出して命令したり，会話したりしているのを見かけるとしたら，それは奇異な印象を受けるであろうし，また周囲に人がいれば混乱や誤解の原因となる．誰に向かって，または何に向かって話しかけているかわからないからである．

近くの人にも，遠くの人にも，また機械にも，等しく有声音声という空気伝達メディアを使ってコミュニケーションしようという考え方そのものに無理があるのではないだろうか．前述の逆説的表現は，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持った人々には音声認識入力は使えない．

個人端末のウェアラブル化，日常生活へのコンピューターやロボットの浸透，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 で描かれる通り，ロボットやコンピューターのマイクロフォンにじかに話しかけることを想定したものであったし，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，音声認識の当たり前の大前提として「空気中に放散された気

導音声を常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インフラ状況では，空中放散音声を相手にするのではなく，むしろ身に着けた携帯情報端末で，「何を言おうとしているか」を認識し，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電気的に相手端末や機械に送った方がはるかに現実的ではないだろうか．しかし，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ある．また長年培ってきた音声認識のすばらしい技術も生かしたい．

2.2 NAM の発見

声帯は振動させずとも，通常音声を発声するときと同じように発話器官を動かすことはできる．ゆるく呼気を吐けば「微弱なささやき声」となるが，これを近くにいる人に聞こえないように行うことは可能である．静かな環境であれば骨導音により，口を動かしている自分にはかすかに聞こえる．

医師は日常的に医療用聴診器で，心音や呼吸音などを聞いている．これは筋肉，結合組織，皮膚，粘膜などの軟部組織，いわゆる肉を伝導する音を，皮膚表面に聴診器という些か歴史的なセンサーを当てて聞いているのである．体幹部に当てることが多いが，これを首や顎の肉の部分に当てて，この「ひそかな発話」を聞いてみると，音質は悪くこもった感じはするが，ほぼ聞き取れることがわかった．自分ではなく他人に当ててもかなりの割合で聞き取り可能である．これは考えてみれば当たり前のことであるが，呼吸音を肉伝導で聞く目的で作られた聴診器で，頭頸部の音響管内の呼吸音を聞いているわけであり，それも「何かを言おうとして調音された呼気音の肉伝導」を聞いているのである．聴診器で肉導音として聞き取れる調音呼気音が，聴診器をはずすと気導音としては全く聞こえない．この聴診器で聞いている対象である「何かを言おうとして調音された呼気音の肉伝導」こそが，伝導媒体まで含めた Non-Audible Murmur（NAM）の概念と定義の由来である．

2.3　非可聴つぶやき認識（NAM 認識）の概念

図 2.1 は．いわゆる無音声認識の一種として考えられる，非可聴つぶやき認識（NAM 認識）入力インターフェースの概念図である．皮膚表面に密着した NAM マイクロフォンにより採音された NAM 音信号は，有線，もしくは無線にてマイクアンプや AD 変換器に送られ，実時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（Hidden Markov Model：以下 HMM）による分析を行う[64]．この場合，認識エンジンは「言語モデル」と「音響モデル」の両方を使用して認識を行う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 音のサンプルを用いた「NAM 音響モデル」を作成して置き換える．こうすればこの音響モデル置き換えの操作以外は，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．

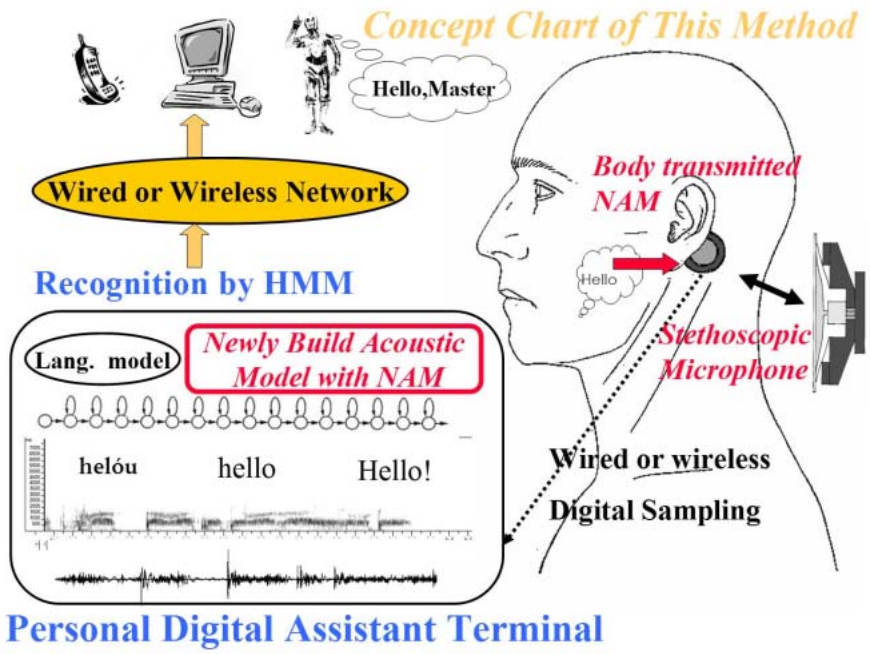


図 2.1　非可聴つぶやき認識の概念図

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．

2.4　体表接着聴診器型マイクロフォンの開発

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 マイクロフォンは当初，医療用聴診器を切断して，中にコンデンサマイクロフォンを埋め込むことから始まった. 図 2.2 は音響モデルを実際に作成した NAM マイクロフォンのモデルである．市販の粘着面のある吸盤用ポリエステル固定板（40mm 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，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狙った．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，低コストで非常に軽量である．また装着感は，馴れれば音楽用ヘッドフォ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.

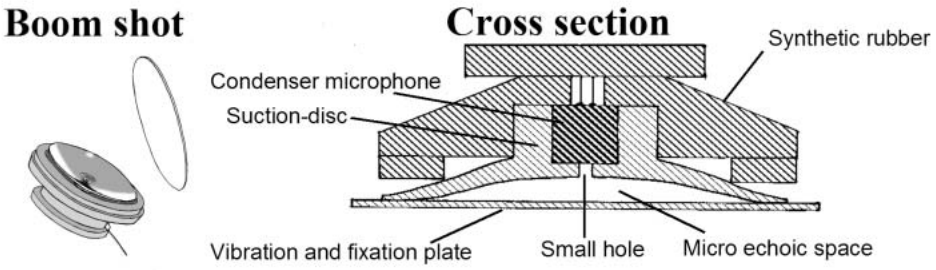
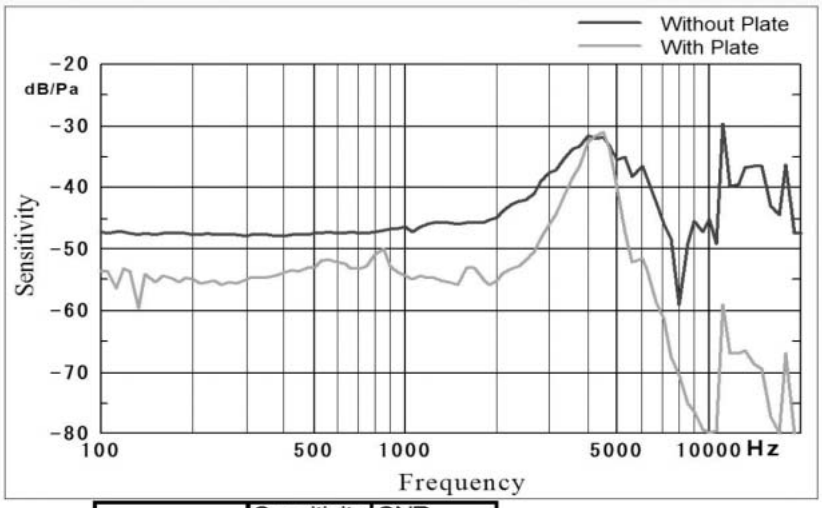


図 2.2　聴診器型 NAM マイクロフォン



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

図 2.3　NAM マイクロフォンの気導周波数特性（非装着時）

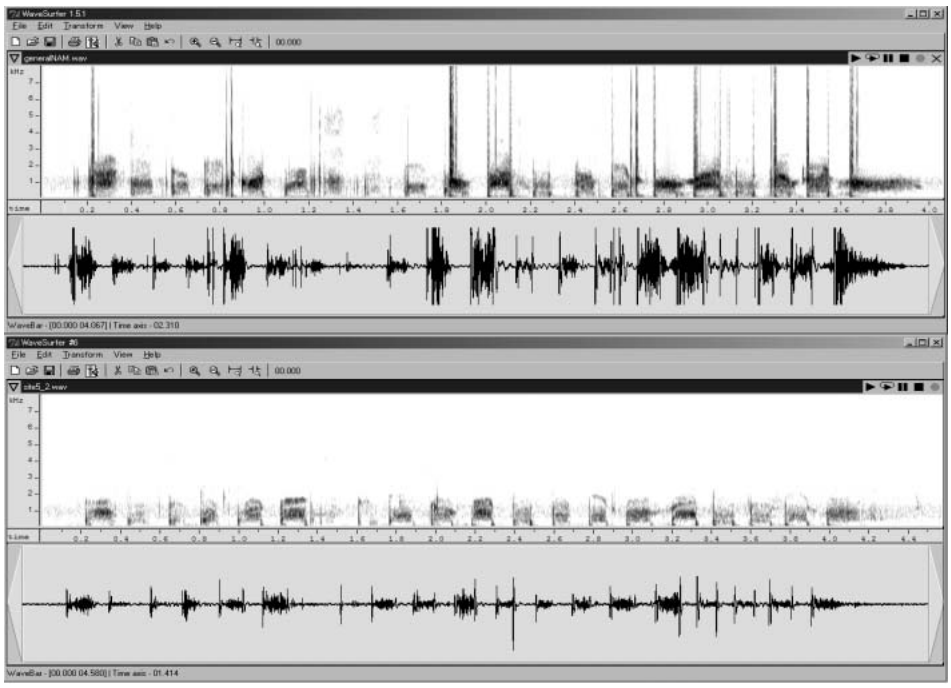


図 2.4　一般的な NAM の音声波形とスペクトラム

図 2.3 に NAM マイクロフォンの気導周波数特性を掲げておく．4〜6KHz で 15dB 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持つ吸盤によって図 2.2 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 2.4 はこの開発した聴診器型 NAM マイクロフォンを，左下顎の耳下腺付近に接着して 16KHz，16bit サンプリングした NAM の音声波形とスペクトラムであり，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．上段ではマイクアンプの出力レベルを上げて，下段では出力レベルを下げて収録したサンプルである．

2.5　NAM マイクロフォン最適接着位置の発見

上記 NAM マイクロフォ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 2.4 の上段の様に子音の音声信号がオーバーフローしてしまい, 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 2.4 の下段の様に，破裂音などの子音を中心に考えてマイクアンプの出力レベルを下げると，母音は小さすぎて不明瞭となるジレンマに陥った．これは NAM マイクロフォンが人間の肉の振動を直接拾うため，肉どうし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 2.5 の左図の番号のごとく頭部の様々な部位に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，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 2.5 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ラムが得られた．

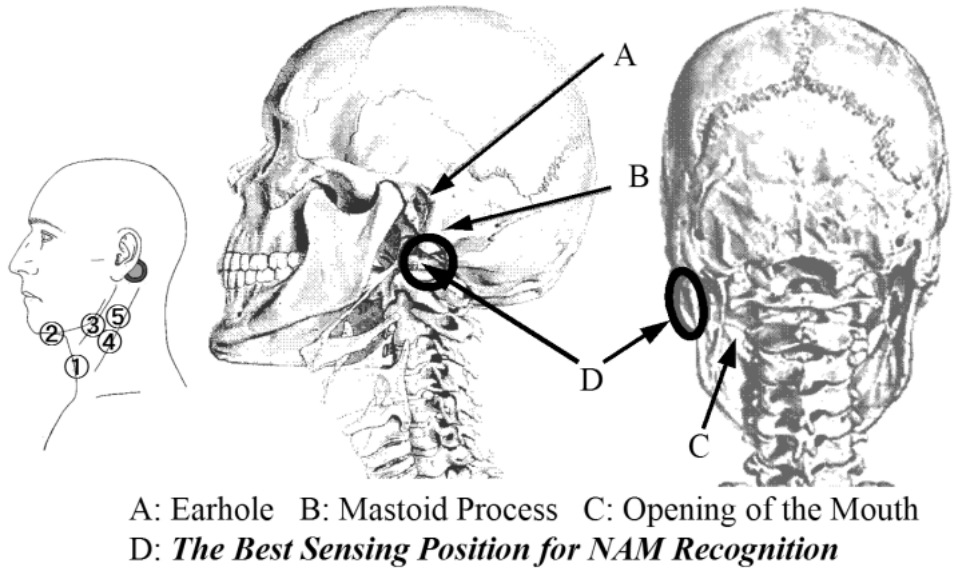


図 2.5　NAM マイクロフォン接着位置

図 2.5 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・母音パワー比が近づき，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，大血管の拍動などによる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剥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，図 2.5 の一番右の解剖図で後ろから口の

開口部が観察できることから，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捉えるに適している．しかも数種類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.

しかしこれより上の，頭蓋骨に当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 S/N 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」の終点に当たる．また最近流行の耳掛け式ヘッドフォ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 マイクロフォン以外に，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フォ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フォンでは，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あまりにも S/N 比が低く使用できなかった.

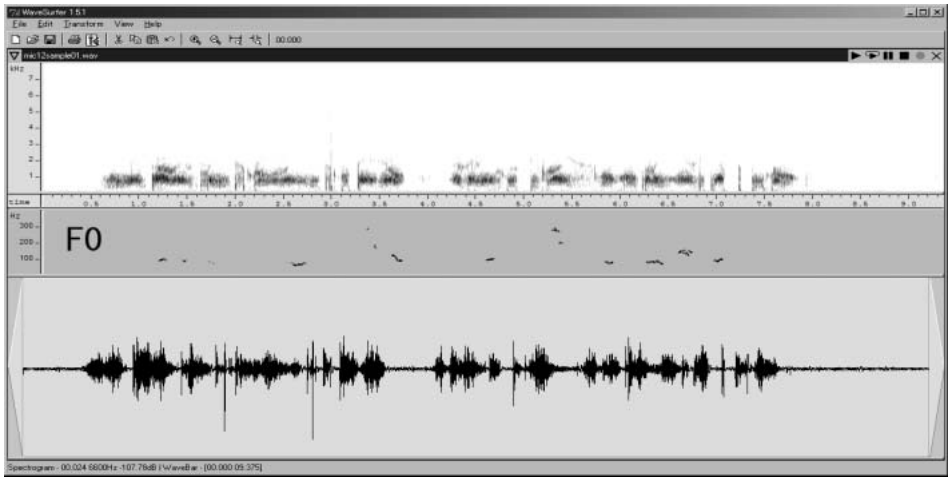


図 2.6　最適位置からサンプリングした NAM

図 2.6 に上記最適位置から NAM マイクロフォンにてサンプリングした NAM のスペクトラム，基本周波数 F0，音声信号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わかる．他部位から採取した図 2.4 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。

2.6 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 2.6 のごとく記録された NAM であれば，内容は慣れればほぼ聞き取ることが可能である．マイクロフォン特性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 マイクロフォ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 5cm 離して空中伝導音を捉えたものと，最適接着位置に装着して肉伝導音を捉えたものを，音声波形，スペクトラム表示し，並列して図 2.7 と図 2.8 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」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 5cm 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，強いて言えば，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 2.7 の右一段目のように，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，声門裂およ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.

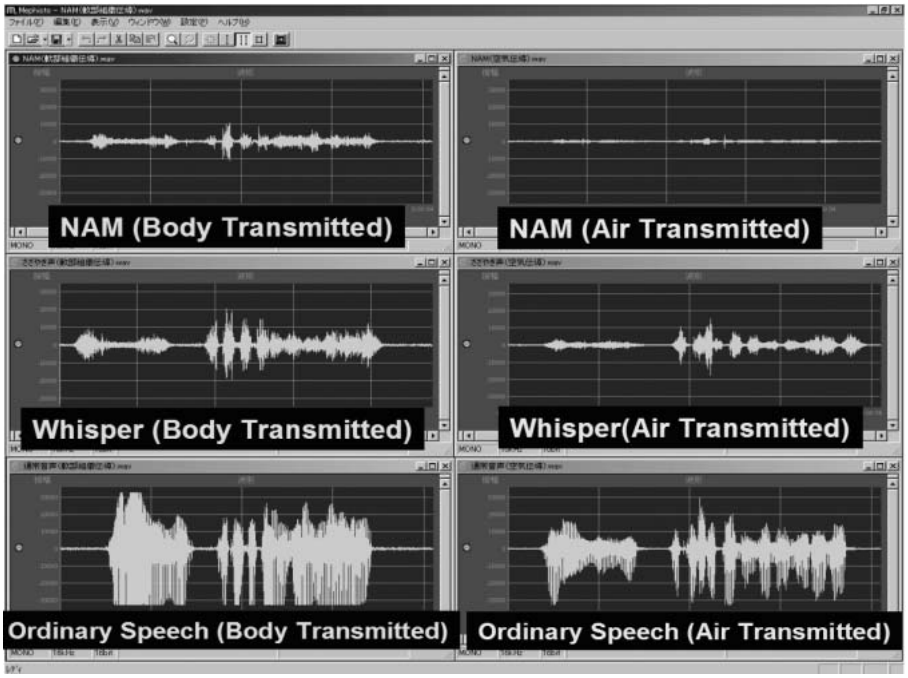


図 2.7　NAM，ささやき声，通常音声の音声波形

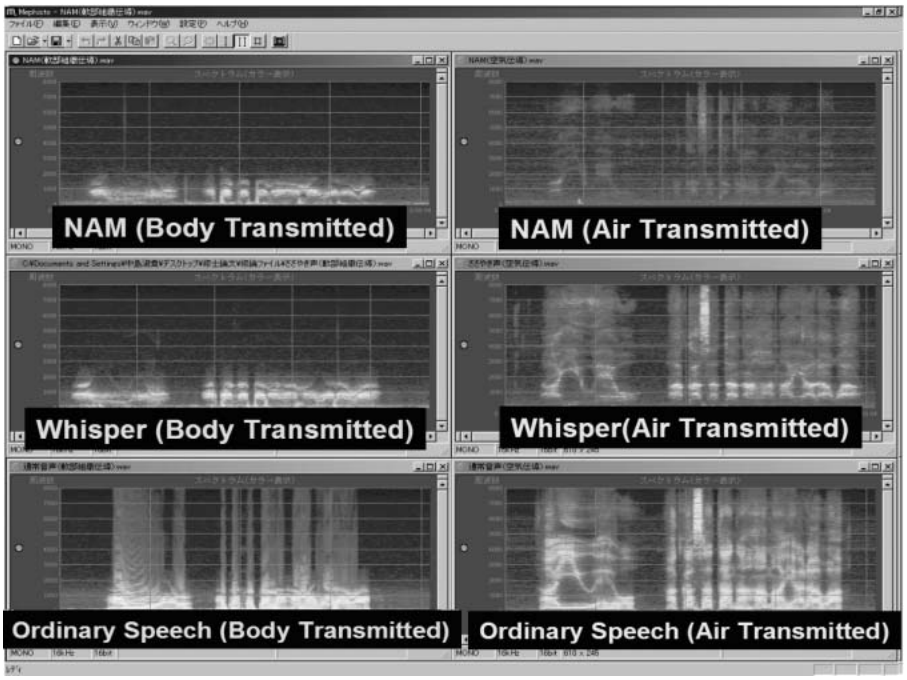


図 2.8　NAM,ささやき声，通常音声のスペクトラム

聴診器型 NAM マイクロフォンでは NAM のスペクトラム描出の際，約 2KHz 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 2.3 の NAM マイクロフォン特性に見られるごとく，むしろ 2KHz～4KHz 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，信号はローパスフィルターを通過した状態となることと，NAM マイクロフォ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情報が少ないためであると考えた．

2.7 NAM の音響モデル作成

NAM を認識可能な HMM 音響モデルを作成するには, 三つの方法がある．

- 初めから NAM サンプルのみを用いて NAM 専用の音響モデルを作る．
- 通常音声の不特定話者モデルに多数の NAM サンプルを EM 学習する．
- 通常音声の不特定話者モデルに少数の NAM サンプルで MLLR などの手法により話者適応する．

上の手法ほどモデルは NAM に特化されたものとなるが，データ収集や音響モデル作成操作が煩雑になる．通常音声の不特定話者音響モデルは既存のものが手軽に手に入るし，NAM サンプル数は下の手法ほど少なくて済む．しかし 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，そのスペクトラムの様相があまりにも通常音声やささやき声とは異なる印象を受けた．したがって初めての NAM 音響モデル作成の試みとしては，できるだけ「純粋 NAM 音響モデル」に近いものを作りたかった．そのため，まず妥協案として，最初に上述の二番目の方法である「通常音声不特定話者モデルに EM 学習を行う方法」をとった．その後に三番目の通常音声不特定話者 PTM モデルに NAM サンプルを話者適応させる方法を試みた．

2.7.1 NAM サンプルのモノフォンモデル EM 学習

ケンブリッジの HMM ツール集である HTK[64]と, IPA の日本語ディクテーション基本ソフトウェア（Japanese Dictation Tool Kit：以下 JDTK）[19]を用いてこれを行った.

学習サンプル文は，特定男性一名の NAM 発話にて，図 2.2 の NAM マイクロフォンを図 2.5 の最適位置（左側）に接着し，室内静環境で NAM 発声にて読み上げた.

読み上げに用いた文章は，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.

マイクアンプは増幅率（電圧利得）26dB のものを用い，計算機は linux+ALSA ドライバの環境で，サンプリング周波数は 16kHz，16bit にて合計 4560 個の NAM 発話による文章読み上げサンプルを収録した.

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）＋ΔMFCC＋ΔLogPow(計 25 次元)にて，Hcopy[64]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[64]にて JDTK の CD-ROM 付属の通常音声の monophone 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 20 回 EM 学習を行った.

認識エンジンは Julius3.4[27]を用い, 音響モデルを変更する以外の条件は通常音声の一般的な認識と同じとし，システムの設定などもデフォルトのままとし，特にパラメータを変更しなかった．言語モデルとしては，JDTK の CD-ROM 付属の 20K 辞書を使用した.

認識率の評価は，JDTK の CD-ROM 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，様々な雑音環境で NAM 発声により読み上げ，これを同じ NAM マイクロフォンにて収録した．この CD-ROM 付属の mkhyp.pl，align.pl，score.pl の 3 つの Perl スクリプトを用いて認識率を集計した．結果が表 2.1 である.

表 2.1　NAM の大語彙連続認識実験（モノフォン EM 学習）

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)

なおテストセット 24 文章の収録環境の内訳は以下の通り.

A：鉄筋のマンション内の静音環境.

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，monophone 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 マイクロフォンでは，側背部の防音が完全でないのと，コンデンサマイクロフォ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．その他に人間の体自体を伝達する外部雑音もある.

図 2.9 に A の静音環境での NAM サンプル数と学習回数による認識精度の上昇をグラフ化したものを掲示する．EM 学習は 10〜15 回程度で飽和することがわかる．またモノフォンモデルでは混合数において 16 から 32 に増加させても認識率に大きな差異は見られなかった.

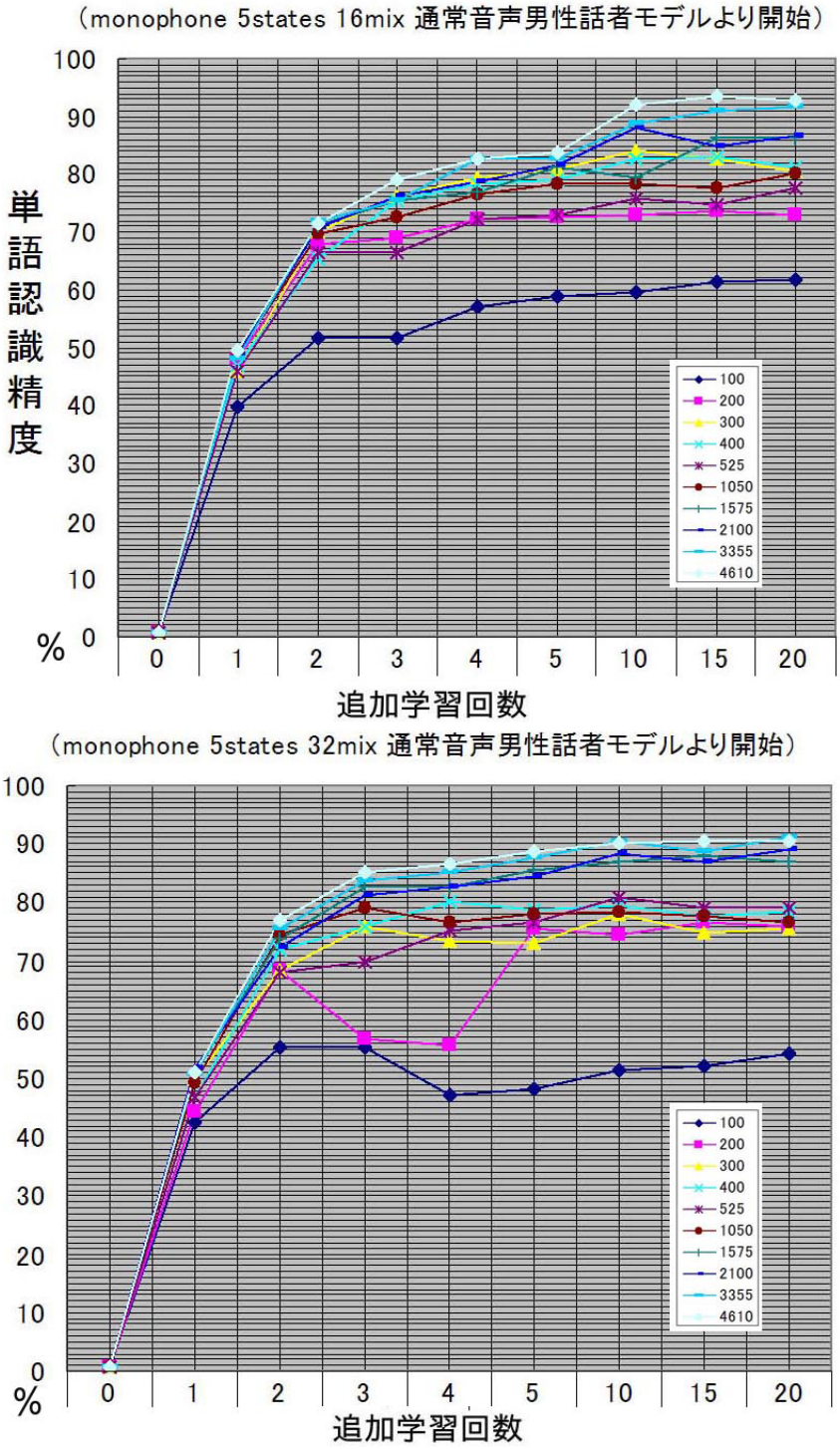


図 2.9　EM 学習のサンプル数と学習回数による認識精度の上昇

2.7.2　NAM と BTOS のモノフォンモデル同時 EM 学習

NAM 認識という非可聴の発話認識を可能にし，認識入力インターフェースとして実現させるためには，その音響モデルは「NAM 音響モデル」を作成する．これにより音声認識入力の便利さや長所を生かしたまま,「入力内容が周囲に聞こえないこと」や「周囲雑音への頑強性」という NAM 認識の大きな利点を享受できる．我々が上記の実験で作成した NAM 音響モデルは，通常音声のモノフォン音響モデルに，多数の NAM サンプルのみを EM 学習したものであるが，この NAM 音響モデルは通常音声では全く使用できなくなり，NAM 専用の音響モデルとなる．しかしそのインターフェースに，時と場合に応じて通常音声でも入力可能であれば，さらに便利であるし応用範囲がもっと広がると考えられる. 自分が何を入力したかを, むしろ周囲の人々に知ってもらいたいときや，自動車内や個室などにいて，周囲をはばかる必要がないときなどがそれに当たる.

我々は，音源が声帯の振動音であるか乱流雑音であるかという違いを除けば，特に認識に重要な低周波域においての（特に第 2 フォルマントまでの）NAM と BTOS のスペクトル包絡構造の概形に極端な差はないと考えた.
音響モデルの混合正規分布の数がある程度あれば，NAM サンプルと BTOS サンプルの同時連結学習により，NAM の分布と BTOS の分布に分離して，また無声子音などは混在して，どちらについても認識可能な音響モデルが作成可能ではないかと考えた.

用いた読み上げ文章は, NAM については ATR 音素バランス文 525 文を 4 回と JNAS の毎日新聞記事 1255 文の計 4610 サンプルであり，BTOS については ATR 音素バランス文 525 文を 2 回の計 1050 サンプルである. サンプリング環境，パラメータ抽出などは 2.7.1 と同様に行い HERest にてモノフォン男性不特定話者モデル（32 混合正規分布）に NAM+BTOS 混合サンプルの EM 学習を 20 回行った．認識率の計算方法は 2.7.1 と同様に 24 文ずつを用い，NAM と BTOS のそれぞれについて計算した.

表 2.2　NAM＋BTOS 同時 EM 学習モデルの大語彙連続認識実験

Env.	Snt	Corr	Acc	Sub	Del	Ins	Err	S.Err
BTOS	24	85.6	82.9	11.9	2.49	2.76	17.1	79.2
NAM	24	87.8	85.3	9.17	3.06	2.50	14.7	79.2
Avg.	24	86.7	84.1	10.5	2.77	2.63	15.9	79.2

(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)

NAM 自身の単語認識精度は NAM 専用モデルより落ちるものの，BTOS, NAM のそれぞれについて，ともに 80%を越え，混在音響モデルでの実用の可能性が示唆された．認識率を下げる一つの要因として，大きめの通常音声では BTOS の信号が一部オーバーフローしてしまったこと，また BTOS のオーバーフローを防ぐためにマイクアンプの出力レベルを下げすぎ．テストセットの NAM 録音の音量が小さすぎたことなどが挙げられる．

この NAM＋BTOS 共用音響モデルによって，NAM と BTOS を混在させて文章を読み上げたものを単独ファイルにサンプリングしたものでも，BTOS 発話部分，NAM 発話部分に関わりなく一文として認識することが可能であった[1]．

しかし学習サンプルの NAM や BTOS とは音量を変えた，小さすぎる通常音声や，大きすぎる NAM など，両者の中間に位置するような移行的な音声は認識が難しかった．

2.7.3　PTM モデルへの話者適応（Iterative MLLR）

前述の通り，HMM を用いて実用的な NAM 認識を実現するためには，本来ならばまず多数話者の NAM サンプルのみを用いて不特定話者 NAM 音響

[1] 詳しくは参考 URL http://www.aist-nara.ac.jp/~yoshi-n/NAM/のデモビデオ「NAM と体内伝導通常音声（BTOS）の単独音響モデルでの同時認識のデモ」をご覧頂きたい．

モデルを作成するべきである．通常音声認識に置いては PTM モデル (Phonetic Tied Mixture Model)という効率の高い音響モデルがあるが，NAM サンプルのみでこれを作成するためには，まずトライフォンモデルを作成せねばならず，数万オーダーの NAM 学習データが必要とされる．しかし NAM 認識においては個人が個人用の NAM マイクロフォンにて認識を行う特殊性から，通常音声の不特定話者 PTM モデルに NAM サンプルを用いて話者適応を行い特定話者 NAM 音響モデルとして使用するという方法がある．この方法のは NAM 学習サンプルの数がはるかに少なくてすむ.

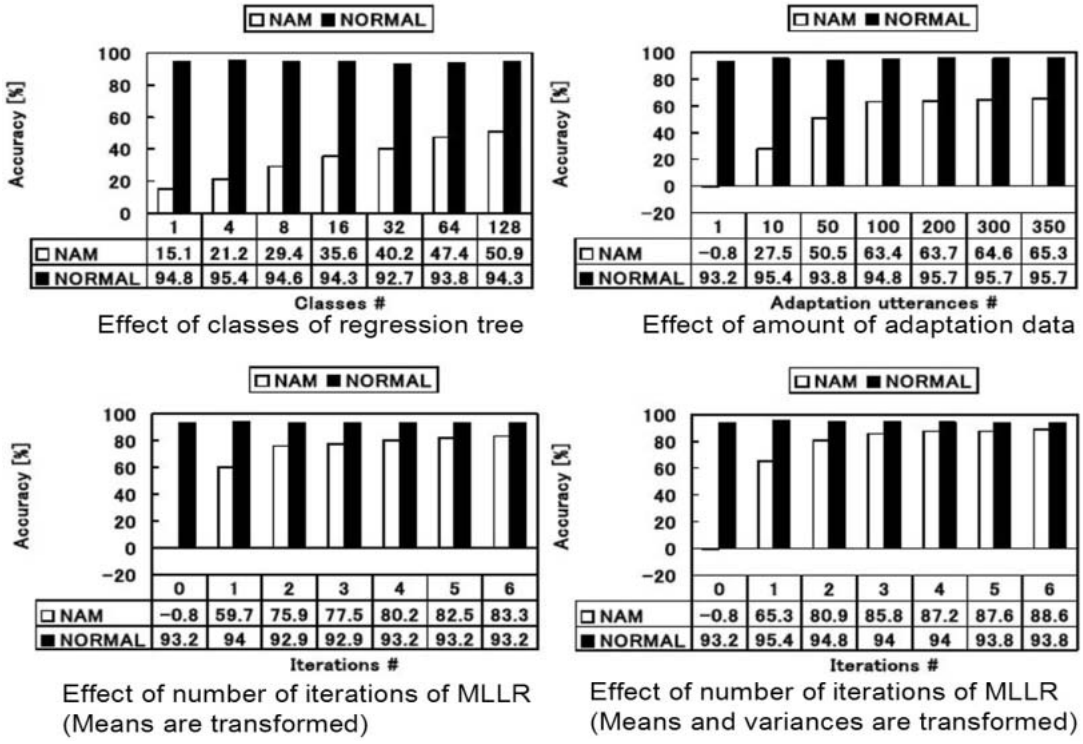


図 2.10　Iterative MLLR による NAM 音響モデルの認識率

初期モデルとして, 3000 状態の通常音声不特定話者 PTM モデルを使用し，話者適応としては MLLR（Maximum likelihood liner regression）[25]を選択し，適応の終わったモデルに対して次々に同じ適応を繰り返す Iterative

MLLR[62]を使用した Panikos らの研究[8]によると，図 2.10 のような結果が得られている．この際話者適応に使用したデータや，評価に用いた 72 個の NAM サンプルは，2.7.1 でモノフォンモデルを作成するときに聴診器型 NAM マイクロフォンで収録した 4610 個の NAM サンプルの中から選んで用いている．

通常音声の話者適応との比較がなされているが，通常音声データは JNAS データベースのものを使用している．この結果によると通常音声不特定話者 PTM モデルへの繰り返し MLLR は 128 クラスターで約 350 文章を 6 回程度行えば，ほぼ最も高い認識精度を得られることがわかる．

また同じ 2.7.1 の NAM サンプルを話者適応と評価に使用し，話者適応に MAP（Maximum A Posteriori）を用いたモデルや，MAP と MLLR を両方繰り返し用いた NAM 音響モデルも作成されており[10]，それぞれ認識精度は 92%を超えている．

聴診器型 NAM マイクロフォンから収録した 2.7.2 の BTOS データを元に話者適応（Iterative MLLR）を行い，同一話者のヘッドセットマイクロフォンで得られた気導通常音声と認識精度を比較した研究もある[13]．表 2.1 の雑音環境分類の A,B,C と同じ環境下で BTOS を収録している．NAM マイクロフォンから得られた BTOS を用いると，ヘッドセットによる通常音声認識より，家庭やオフィスレベルの雑音環境下の認識に頑強であることがわかる．

表 2.3　BTOS とヘッドセットマイク収録通常音声との認識精度の比較

Word Accuracy [%]			
Microphone	Environment		
	Quiet	TV-news	Music
Close-talking	94.4	91.7	91.9
BTOS	93.8	93.2	92.9

また NAM の不特定話者モデルを作るベースラインの検討としての研究[14]もあるが，聴診器型 NAM マイクロフォンの帯域が狭く，またサンプリ

ング技術もデータ収録当時は未熟であったため，あまり良質な NAM データが集積されていない．21 人（男性 14 人：女性 11 人）の 3189 個の NAM データで通常音声不特定話者モデルに話者適応を用いている．

表 2.4　聴診器型 NAM マイクロフォンによる NAM の不特定話者モデル

Word Accuracy [%]				
Speaker	Training Technique			
	MLLR	MAP	MLLR+MAP	EM
Male	73.5	62.4	77.8	75.9
Female	70.0	58.4	71.1	65.4
Average	71.8	60.4	74.4	70.6

今後, 計画的かつ大規模に NAM データ収録を行い, 純粋に高音質な NAM データのみから作成された NAM 不特定話者 PTM モデルを作成すること，また BTOS も同時に収録し，NAM と BTOS を同時に認識する，総合肉伝導音声不特定話者 PTM モデルの作成が課題である．また音声データのパラメータ化も NAM や BTOS の認識に効果的な抽出法があるはずであり, 気導通常音声認識との認識率の本当の比較は，それらの確立後に待たれる．

2.8　まとめ

音声認識入力が内包する日常的普及への本質的欠点に気づき, いわゆる「無音声認識」の実用的価値と，通常音声ばかりでなく NAM を認識に用いることの必要性を考察し，NAM 発見の経緯を述べた．聴診器型マイクロフォン開発，最適装着位置の発見，NAM 音響モデル作成により，大語彙連続認識の実験結果より NAM 認識の可能性を論じた. NAM 認識の大きな特徴は「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．総合発話認識入力としての応用もめざし，BTOS も同時に認識に使用できることを検証した．

この認識入力方式は，通常音声認識と比較して，通常音声で発話できない

ハンディキャップを持った人々を支援する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時，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．それは音声認識の長く，たゆまない技術蓄積のもとにはじ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素晴らしい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．

図 2.11　聴診器型 NAM マイクロフォンの外観

第3章 ソフトシリコーン伝導型 NAM マイクロフォン

3.1 はじめに

それはまだ黒い電話が台所にあったころ．夕食の団らん時にベルがなる．電話をとった父親が「女の子からだぞ」と受話器を無造作に渡す．背中に家族の視線を受けながら「ああ」「うん」「いや」「まあ」「いいよ」「じゃあね」などと，無愛想で無難な受け答えを，小声でした経験はないだろうか．

人類が音声言語によるコミュニケーションを始めてから何万年になるのかわからない．しかし音声言語を電気信号に変換し，それをモールス信号によって普及しつつあった電信網というインフラにのせて，遠隔地にいる人と会話をするようになってから 120 年程度である．

声帯を振動させて音源とし，空気を介して伝達する通常の音声は，音量の調節により隣の相手から数十メートル離れた相手にまで情報を伝達することが可能である．しかしその相手までの距離を半径とする円形の範囲（球形の範囲）にいる第三者にも，同時にその情報は伝達しうる．場合によっては，その範囲外の人々にも伝わる．その第三者にとっては，その情報が必要なこともあるが，むしろ迷惑な場合もある．また話者にとっても，その情報が第三者にも知れても構わない場合もあるであろうし，また困る場合もある．

電話は距離を限定しない特定の相手だけに音声言語による情報を伝達するために，西洋の堅牢な個室型社会で生まれた発明だったと言ってもよい．しかし電話の発明者は，まさか人々が室内を離れてあらゆる場所でコンパクト

サイズの超小型のコンピューターを耳に当てて電話をするなどとは，想像もしなかったであろう． どこでもいつでも会話できるからこそ，通常音声では上記のような問題が起こる場合があり，心おきなく発声できる場所に移動しなければならない．当然電車や映画館内での携帯電話の使用は禁止される．

こういう社会の状況になれば「声を出さずに」電話をかけられるとしたら，つまり NAM による，いわゆる「無音声電話」が実現したとしたら，電話をかける本人にも，また周囲にいる第三者にもありがたいのではないだろうか．

NAM 音は増幅する以外の信号処理を施さずにそのまま聞いても，「音のこもったささやき声」として，聞き慣れればある程度聞き取り可能である．また NAM 音信号に声質変換や音源付与などの信号処理を加えて通常音声化する試みも開始されており[15][58]，これが NAM よりも聴き取りやすい形でリアルタイムに実現すれば，前述の「無音声電話」が実現する．しかも NAM マイクロフォンを使用することによって，ヘッドセットマイクなどから気導音声をサンプリングするよりも，外部雑音に頑強であるため[55]，話者本人が周囲の状況へ気兼ねする必要もなく，また周囲からの雑音の影響も受けにくいという，話者の置かれた環境にきわめて依存しない電話となる．

それを実現するに当たっては，気導音を排して肉伝導音だけをサンプリングするという NAM マイクロフォンの基本発想を崩さずに，サンプリングされるオリジナルの NAM 音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．またそれは NAM 認識においてもその認識率向上の鍵となる．そこで NAM マイクロフォンの設計を根本から見直してみた．

3.2 聴診器型 NAM マイクロフォンの欠点

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 2.2 のごと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: 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．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 以上の高域に急激なカットオフが見られ，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 3.1 の下段は聴診器型 NAM マイクロフォンによる 16KHz サンプリングの NAM 音（左）と BTOS 音（右）のスペクトラムである．上段に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）のスペクトラムを比較のために掲げる．

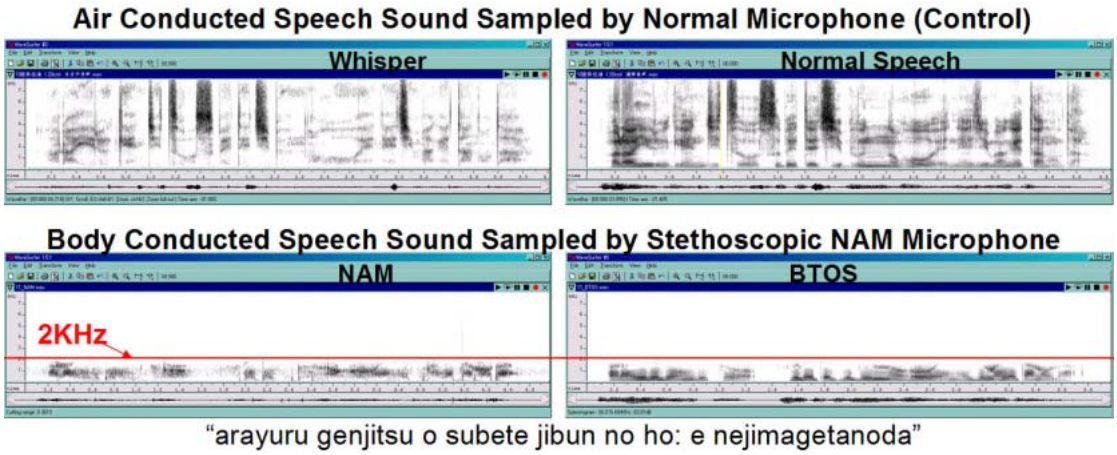


図 3.1　聴診器型 NAM マイクロフォンによる NAM と BTOS
（気導音の通常音声とささやき声との比較）

聴診器型 NAM マイクロフォンでサンプリングした NAM 音は，HMM 認識が可能であるとは言っても，増幅して聞いた印象では「こもったささやき声」に近く，聴き方に習熟するとある程度聞き取り可能であるが，初めて聞く者には発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 NAM 音よりは良いが，「非常に音のこもった通常音声」となる．

3.3　帯域を広範化させるために

元来聴診器は，非常に低域の周波数帯の心音や呼吸音などの体内伝導音を詳細に聴くために作られている．聴診器を模した聴診器型 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，コンデンサマイクロフォンから皮膚までの密閉空気層と固体の振動板が，むしろ帯域を狭小化させている可能性がある．また聴診器型 NAM マイクロフォンはその密閉の完全さが問題となり，微量な空気の漏れがあるだけで，気導外部雑音が容易に大量に漏入する．そしてコンデンサマイクロフォンの振動電極自体は通常通り空気に接するようになっているため，漏入した雑音に敏感に反応する．またもし密閉が完全であったとしても，吸盤の薄い樹脂一枚を通して外気に触れる構造になっているため，吸盤の背面よりの気導外部雑音に脆弱である．

NAM マイクロフォンの元来の，そして究極の目的は，マイク側背部からの漏入・伝導する外部雑音をはじめとする気導音を排して，NAM のごとき肉伝導音だけを広帯域，高感度で拾うことである．気導音の漏入を許せば帯域は伸びるが，外部雑音への感度が気導マイクの感度と変わらなくなる．

そこで振動板とコンデンサマイクロフォンとの間に介在する音媒体としての空気を，完全に排除することを試みた．新たに音媒体として弾性があり，形成が容易で，人体への接触に安全な歯科技工用のシリコーン（ラボシリコーン：松風）を選択した．

またコンデンサマイクロフォンは気導音を採録するために設計されているため，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．ここに硬化前のシリコーンを詰めて NAM マイクロフォンを作成すると，聴診器型では見られなかった 2KHz 以上のフォルマントがわずかに描出されたが，聴診器型に比し感度は著しく低下した．またシリコーンの硬化に伴いシリコーン体積に軽微な縮小がみられ，振動電極板との接触が保てなくなった．

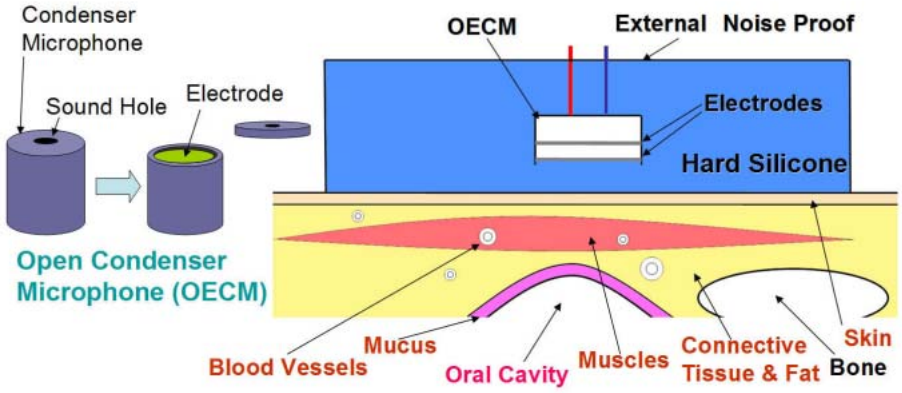


図 3.2　ハードシリコーン型 NAM マイクロフォン
Open Condenser Wrapped with Hard Silicone Type (OCWHS 型)

そこで図 3.2 の左のごとく，伝導媒体を直接，振動電極板の全面に密着させるため，この小孔の開いた表面金属を丁寧に削り取り，振動電極板を完全に露出した形態のコンデンサマイクロフォンを作成した．これを仮に Open Electret Condenser Microphone (以下 OECM)と名付ける．

この OECM を，図 3.2 の右のように硬質の消しゴム～プラスチック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指で圧着してみた．これを Open Condenser Wrapped with Hard Silicone Type (OCWHS 型)の NAM マイクロフォンと呼ぶことにする．

図 3.3　OECM の製作過程

OECM の実際の製作過程を図 3.3 に示す．一番左側が，通常の ECM の感音面とその裏側である．左から二番目が，表面の防塵用の黒い皮膜をはがし

たもので，気導音を通すための小孔が表面に開いている．この表面を丁寧にヤスリで削っていくと，左から三番目のように，周囲の側面金属と前面の金属との連続が途切れる．残って蓋のようになった前面の金属を剥離すると，一番右側のように振動電極板の露出した OECM となる. この工程において，少しでも振動電極に傷をつけると, OECM の感度は著しく低下するので細心の注意をもって剥離する.

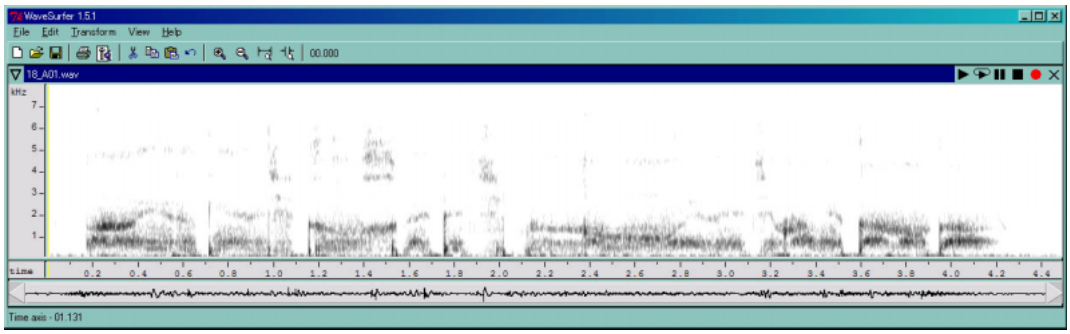


図 3.4　ハードシリコーン型 NAM マイクロフォン（OCWHS 型）でサンプリングした NAM 音のスペクトラム

このハードシリコーン型 NAM マイクロフォンを用いると，聴診器型に比し, 感度は同程度に近づき, 図 3.4 の NAM 音のスペクトラムのごとく 2KHz 以上の帯域も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．図 3.4 の収録内容は聴診器型の図 3.1 と同じく「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.

こういった弾性固体媒体で OECM を包み込んで，その振動電極板の全面に直接接触させることにより，空気との接触がなくなり，振動電極板への気導音の漏入は，完全に排除できる．側背部の外殻から伝導する軽微な外部雑音はこの時点でやむを得ないとして，ほぼ肉伝導音だけをピックアップできる．しかも音声における音韻の識別にとって重要な 2KHz 以上に，帯域が拡大した.

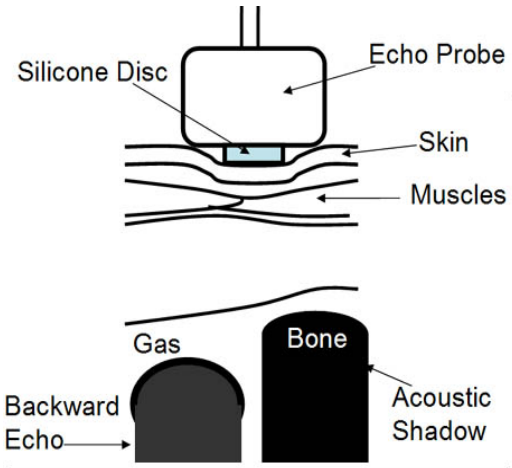
3.4　接触面感度を上昇させるために

それでは何を音媒体に用いれば，最も効率よく肉の振動をコンデンサマイクロフォンの振動電極板に伝えることが可能であろうか.

二つの物質の音響インピーダンスが異なれば異なるだけ，その物質間の接する境界面で強く音の反射が起こる．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, 音のこの性質と体の軟部組織（いわゆる肉）の微妙な音響インピーダンスの差異を用いた音の反射距離計測を利用している．体内構造をこの装置で描出するときに，軟部組織の構造はよく描出されるが，水の音響インピーダンスに近い軟部組織とは明らかに音響インピーダンスの異なる骨やガスなどは，表面で超音波がほとんど反射してしまうため，その表面以降の構造は影となってまったく描出されない.

図 3.5 の左上の模式図のように，超音波イメージング装置の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，観察できる体内イメージを検討してみた．こうすることによって，はさみこんだ物質と人間の軟部組織との音響インピーダンスがどれくらい異なるかが視覚的にわかる.

図 3.5 の 4 枚の写真に見られるように，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プローブと皮膚の間に置いたときには，その境界面で超音波のほとんどすべてが反射を起こし, 超音波がより深部に届かないため, 金属プレート以降の人体内部構造が全く描出されなかった．また前述のハードシリコーンのプレートをはさみこんだ場合，ごくわずかに体内構造が描出され，もともと高輝度のラインのみ描出された．人の軟部組織の柔らかさと弾性に最も近いと感じられるソフトシリコーンのプレートを用いると，プローブと皮膚との間に何もはさまなかった時と同様に，ほぼ完全に体内の構造が描出された．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，肉に似た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

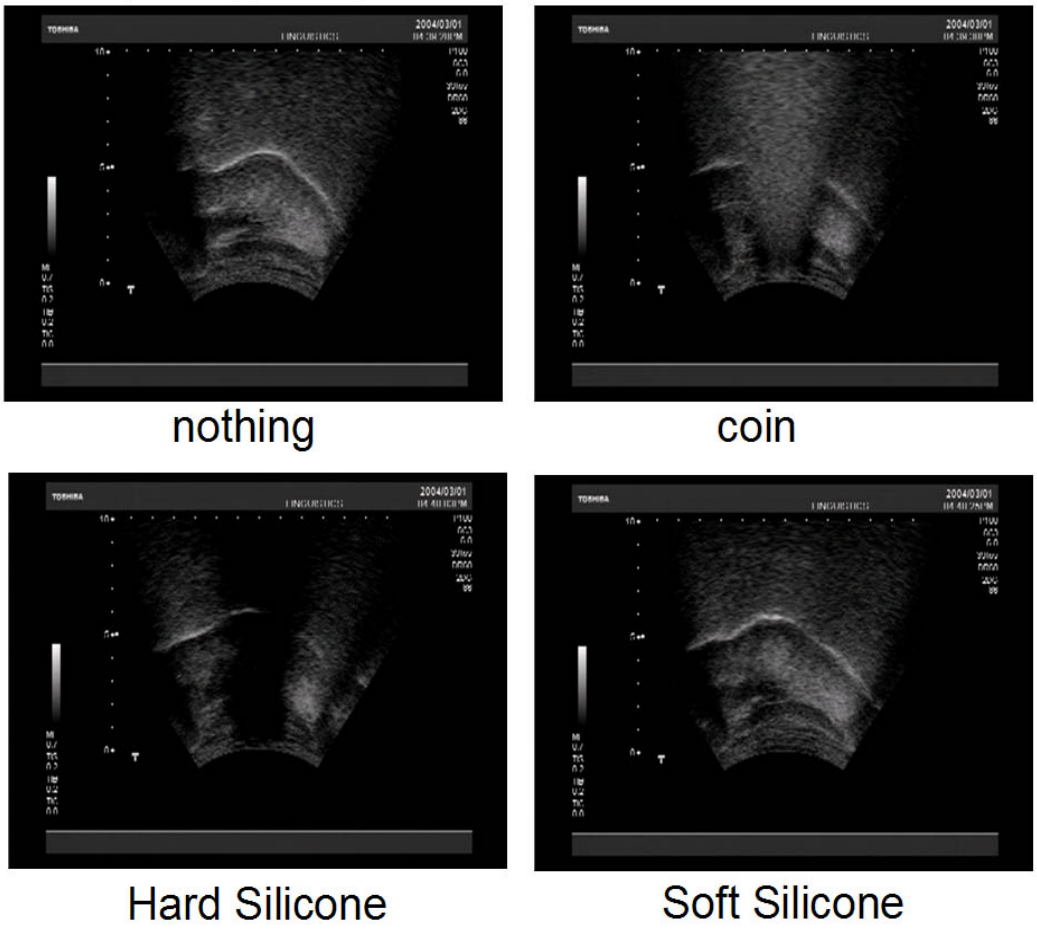


図 3.5　医療用超音波イメージング装置を使って視認できるさまざまな物質の音響インピーダンスと人間の肉の音響インピーダンスとの差異

3.5　ソフトシリコーン型 NAM マイクロフォン

もともとコンデンサマイクロフォンや圧電素子などのセンサーを手術で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（実際には感電する）, 肉伝導音センサーである NAM マイクロフォンの効率としてはそれがベストかも知れない.

しかし人間の肉と同じ音響インピーダンスを持つような軟らかいゲル状のシリコーンを「擬似肉」として皮膚に密着させ，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 の振動電極をその擬似肉に直接接触させたりすれば，上述のような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のではないかと考えた. そこで前の図3.5で，体内構造を観察するのにまったく影響のなかったソフトシリコーンを音媒体に用いて，NAM マイクロフォンを試作した．いろいろなタイプのものを数多く作成してみたが，結局の所，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.

- 第一のタイプは，図 3.6 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ーンに置き換えた形である．OECM の振動電極はまったく空気を介することなしに，ソフトシリコーンだけを媒体としてその全面に肉伝導音を拾うことになる. 図 3.6 の中間色で塗りつぶした部分が音響インピーダンス的には似通った物質ということになり，空気と媒体こそ違え，やや隠喩的表現となるが，音にとっては一種の洞窟と同じ肉伝導の反響空間となる．また視点を変えれば, 直接皮膚に接触させたかった OECM の振動電極を, 感電させることなく, 言わば「擬似肉のコブ表面」に直接接触させた形となっている. しかも聴診器型の構造を踏襲することによりパスカルの原理による感度上昇も期待できる.

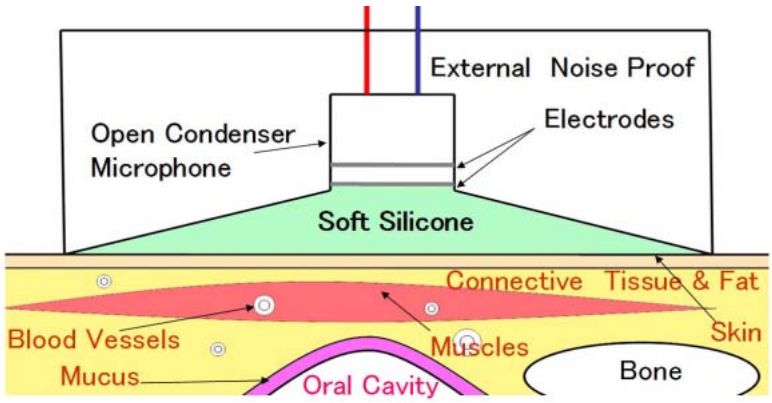


図 3.6　Open Condenser Mediated with Soft Silicone Type (OCMSS 型)
ソフトシリコーン伝導型 NAM マイクロフォン

◆ 第二のタイプは図 3.7 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を音響的に隔離しやすい．またちょうど「擬似肉のコブ」内部に OECM を直接埋め込んだことになり，皮膚に直接埋め込むのと音響的に同様の効果が期待できる．

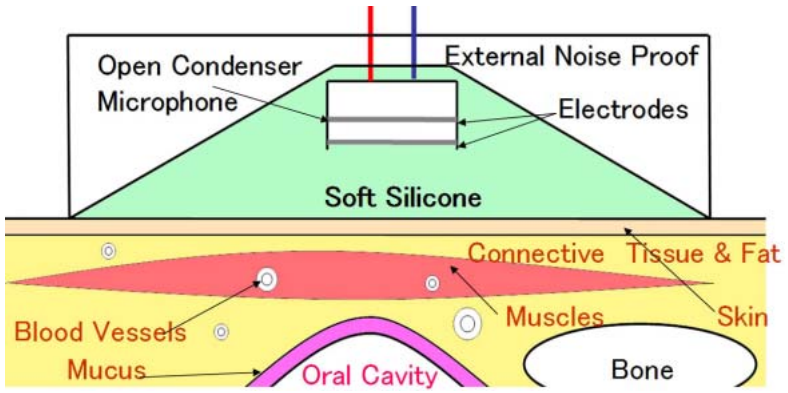


図 3.7　Open Condenser Wrapped with Soft Silicone Type (OCWSS 型)
ソフトシリコーン伝導型 NAM マイクロフォン

◆ 第三のタイプは，OECM のようなコンデンサマイクロフォンではなく，図 3.8 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．

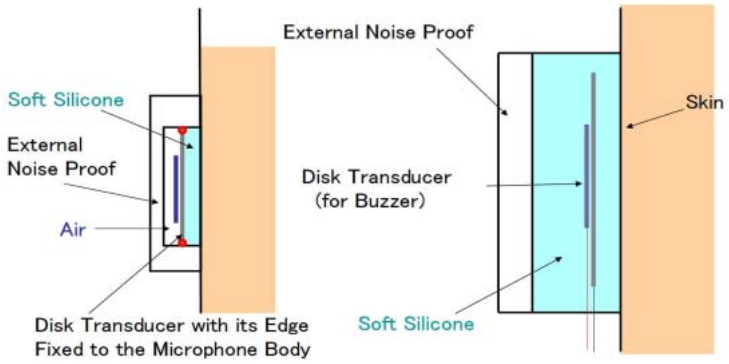


図 3.8　Transducer Mediated with Soft Silicone Type (TMSS 型)
ソフトシリコーン伝導型 NAM マイクロフォン

以上 3 タイプとも，ソフト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，指で聴診器型と同じ位置に圧着して使用した．

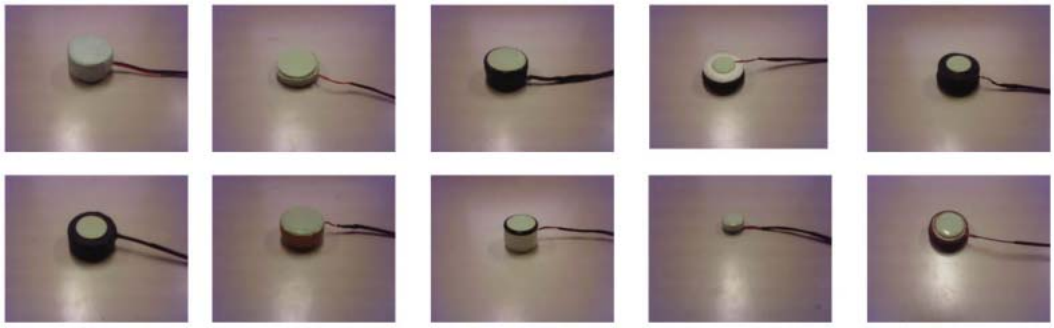


図 3.9　ソフトシリコーン伝導型 NAM マイクロフォン試作品の外観

3.6　NAM マイクロフォンの視覚的簡易評価

気導音をサンプリングする目的の通常マイクロフォンと違って，接触型の体伝導音マイクロフォンの特性を評価するための方法はまだ確立していない．NAM マイクロフォンの接触面を通常の気導マイクロフォンの感音面とみなして，その空気中でのインパルス応答を測ってマイクロフォン特性とするような方法で特性を測ることは，参考資料として必要ではあるが NAM マイクロフォンの目的に添った評価法とは言えないと考える．気導外部雑音を排して体伝導音だけをサンプリングしたいという接触型マイクロフォンの目的から言えば，接触面の振動に対する感度が高ければ高いほど，それ以外の空気に接した部分の気導音に対する感度は低ければ低いほど良い．つまり考えなければならない感度はどうしても二つあることになり，空気媒体の振動から得られる感度がすべてである通常気導音マイクロフォンとは違った評価法があってしかるべきである．二つの NAM マイクロフォンがあったときに，そのどちらを使うべきか判断するための指標が欲しい．本来ならば弾性固体の振動を精密に計測するような標準的な機器があれば望ましく，なんらかの形で数値評価できればいいのだが入手困難であり，またそれを使用しての測定法にもいろいろ議論が分かれる．

そこで二次元視覚情報に頼ることになるが，NAM マイクロフォンの種類別性能の紹介のために，感覚的にその特性がわかりやすい視覚的評価法を考案してみた．これによって自作 NAM マイクロフォンの簡単な性能比較ができる．また骨伝導マイクロフォンや Throat マイクロフォンなどの種々の接触型マイクロフォンも，その目的が同じである限り，同列に評価できる．

接触型体伝導音マイクロフォンを評価する三つの指標を，「帯域」「皮膚接触面感度」「気導外部雑音への頑強性（気導音感度）」とした．また NAM サンプリング目的と異なり，BTOS サンプリング目的では，経験的に増幅率を 20dB 程度落とすことが可能なため，「外部雑音への頑強性（気導音感度）」については NAM 目的と BTOS 目的でそれぞれ別に考える必要がある．

3.6.1　帯域

図 3.10 は種々のタイプの NAM マイクロフォンで 16KHz サンプリングされた NAM 音のスペクトラムである．また図 3.11 は BTOS 音のスペクトラムである．どちらも下の三種が OCMSS 型，OCWSS 型，TMSS 型のソフトシリコーン型 NAM マイクロフォンである．接触面感度向上の目的で音媒体にソフトシリコーンを使用したが，NAM においても BTOS においても，聴診器型だけではなく，ハードシリコーン型よりさらに帯域も高域に広くなっていることがスペクトラムを比較するとよくわかる．

図 3.10 の NAM において OECM を使用した OCMSS 型と OCWSS 型を比較すると，一般的に OCMSS 型の方が帯域は広くなる．「sh」などの高域の摩擦性の子音がどの高さまで描出されるかということ以外にも，共鳴音である母音のフォルマントの描出限界も大切である．作り方によるばらつきはあるが，どちらも母音のフォルマントが 4～4.5KHz 程度まで，高音の子音については OCMSS 型の方がより高域まで描出される傾向にあった．

セラミック型圧電素子を利用した TMSS 型は描出の限界は OCMSS と変わらないほど帯域は広い．これは圧電素子の外縁をプラスチックで固定して皮膚接触する方の片面をソフトシリコーンに，圧電素子の歪みが大きくなるようにもう片面の裏面は密閉空気としたものである．このタイプは特に帯域が広く, OECM を使用した二つのタイプではフォルマント描出の希薄になる 3KHz～5KHz の帯域がより明瞭となる. 逆に 1KHz 以下の低域が淡くなり，より「ささやき声」のフォルマントに近い形となり，聴取した印象も「音のこもり」がなく，より「ささやき声」に近い．BTOS についても傾向は NAM と同様であるが，OCWSS は感度が最も高いため，マイクアンプの増幅率を 0dB 近くに下げないとオーバーフローのためフォルマント描出不能となる．

いずれにせよ NAM，BTOS ともに，ソフトシリコーンを音伝達媒体に用いたソフトシリコーン型の NAM マイクロフォンは，聴診器型に比して高域の子音や母音のフォルマントを描出可能であり，帯域の広範化を認めた．

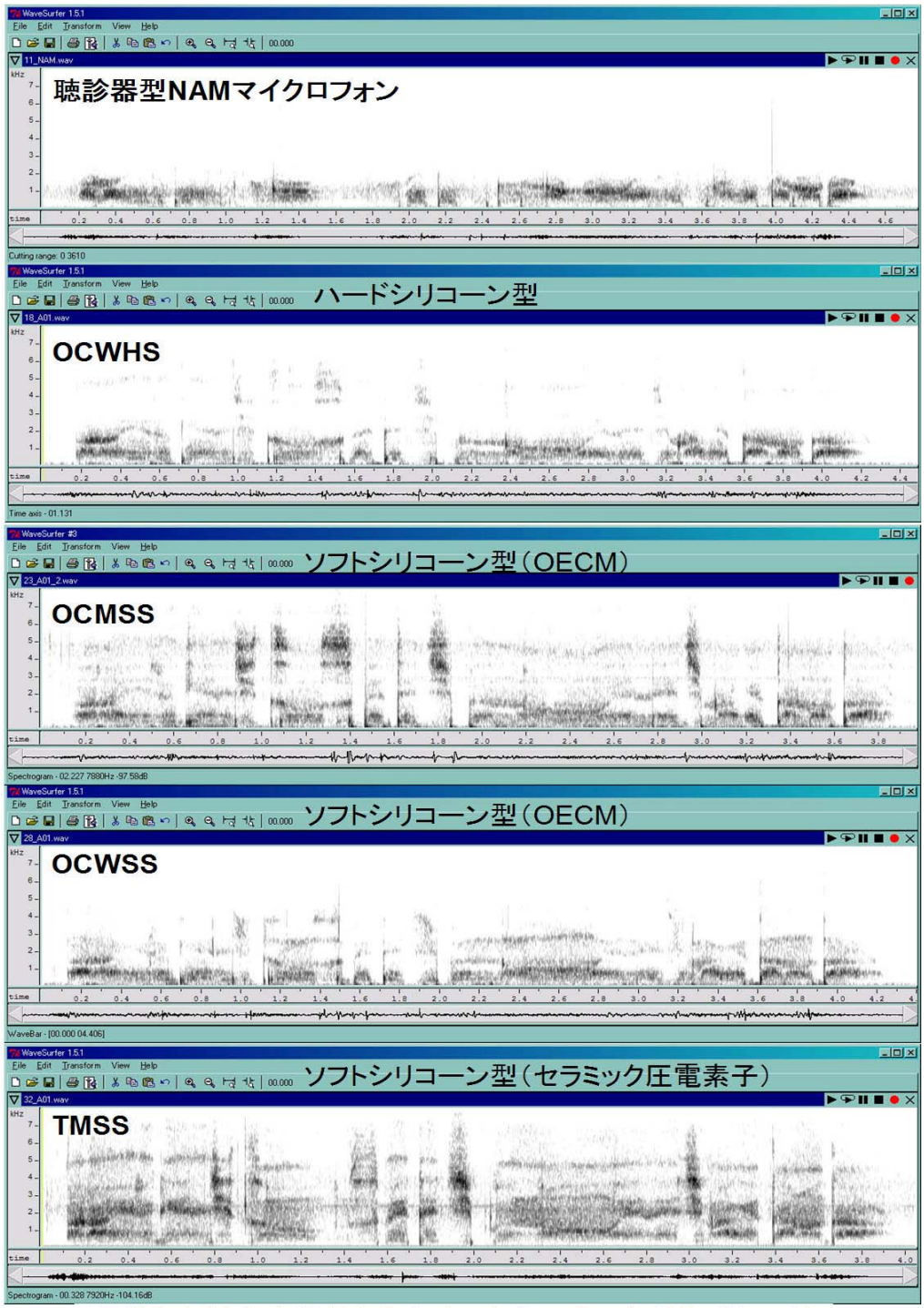
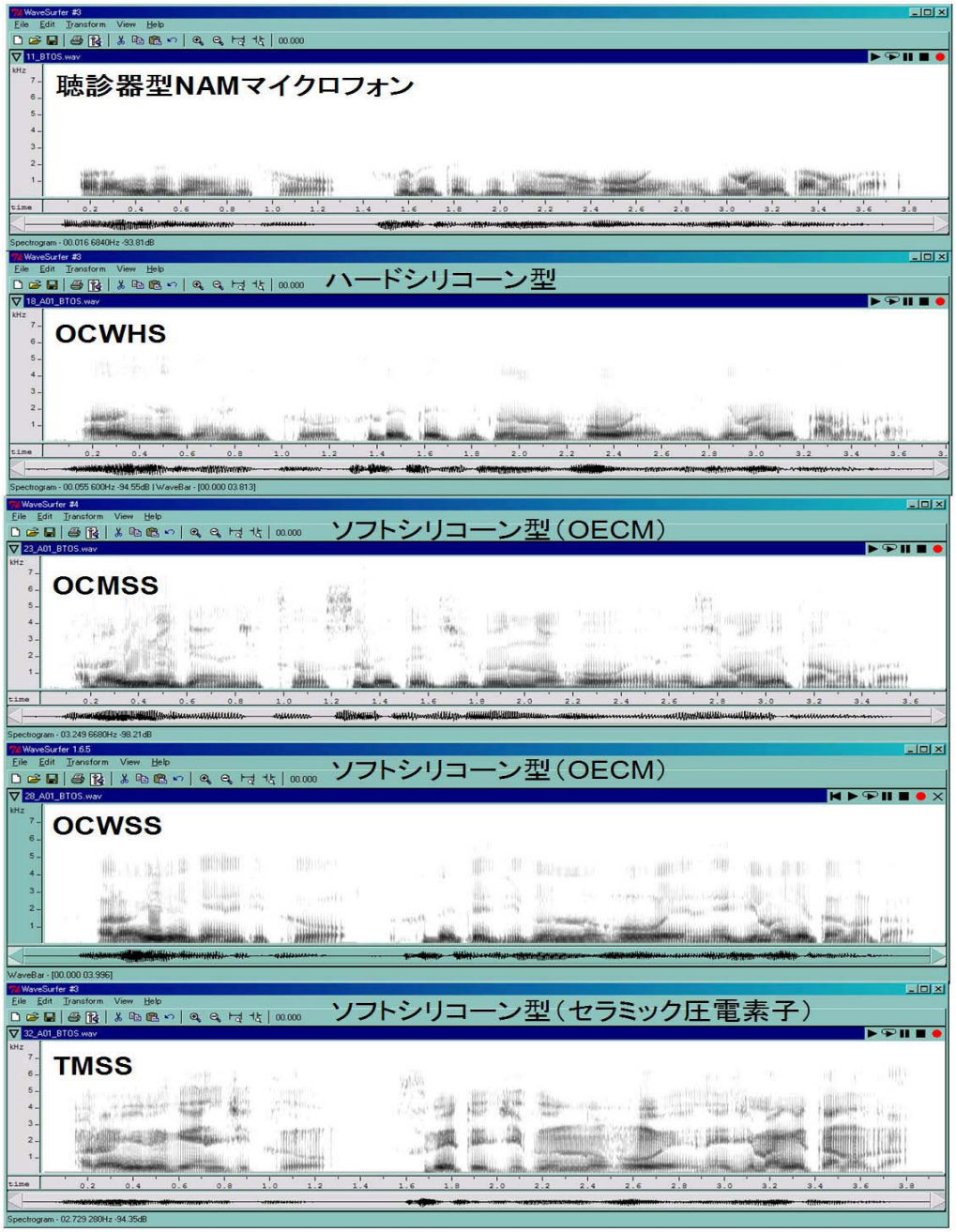


図 3.10　NAM 音スペクトラムによる帯域比較



発話内容「あらゆる現実をすべて自分の方へねじ曲げたのだ」

帯域の広さの順位	1	2	3	4	5
NAM マイクロフォンの型	TMSS	OCMSS	OCWSS	OCWHS	聴診器型

図 3.11　BTOS 音スペクトラムによる帯域比較

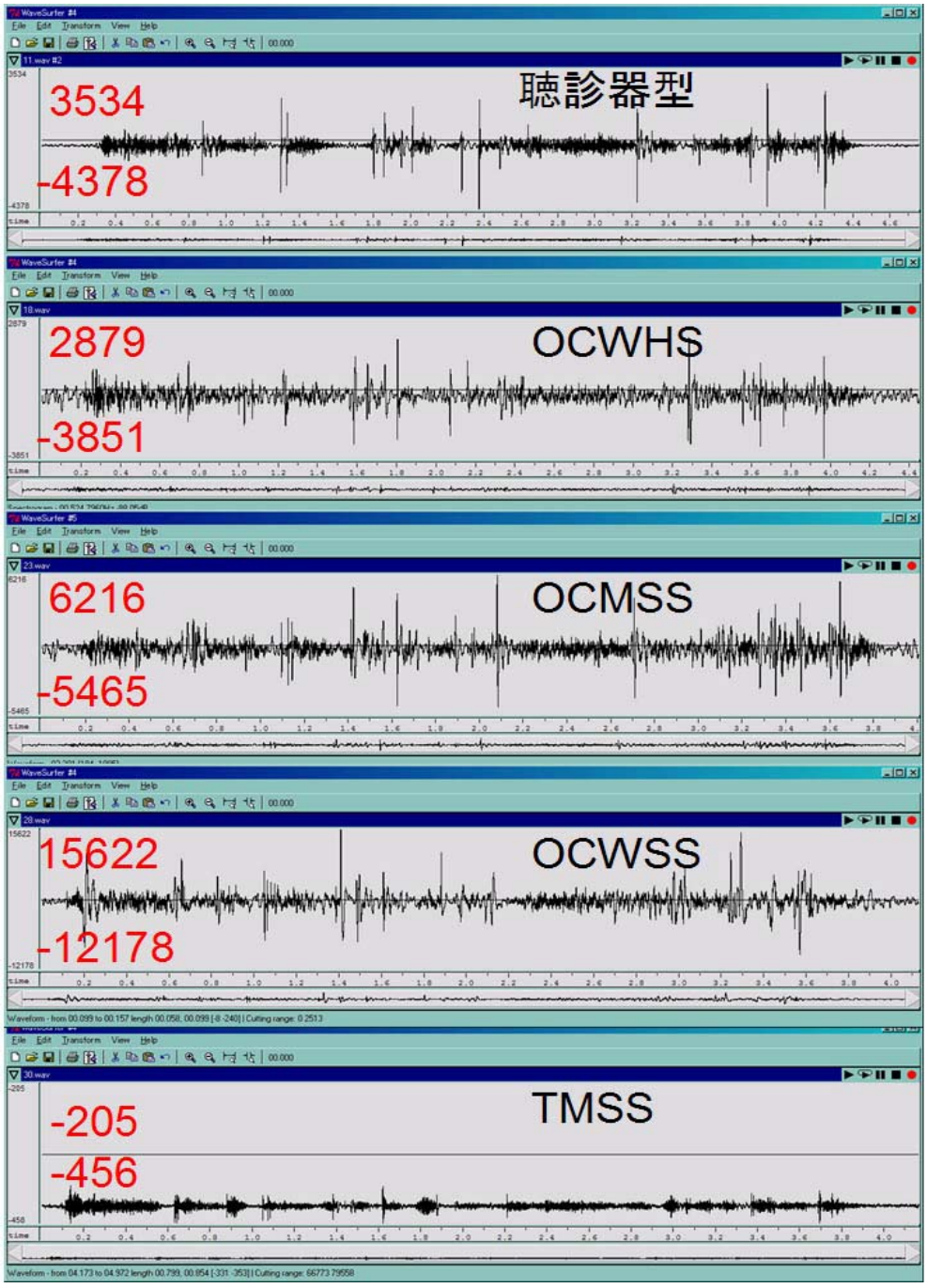
3.6.2 皮膚接触面感度

NAM マイクロフォンの皮膚接触面感度は,マイクアンプの増幅率と出力レベルボリュームを一定にしたときの NAM 音信号の最大振幅で評価することとする．製作した NAM マイクロフォンの接触面感度は，その構造や使ったコンデンサマイクロフォンや圧電素子の種類によって大きく異なる．マイク間のばらつきを勘案して数値的に比べやすいようにこの論文では自作マイクアンプ（電圧利得：26dB）の出力レベルに 50KΩの抵抗をかけた時に得られた信号振幅を用いることにした（もちろん他の抵抗値でも構わない）.

図 3.12 に NAM マイクロフォンのタイプ別皮膚接触感度を掲げる. 左側の数値が振幅の数値である（16 ビットサンプリングのため最大値 32768）.

NAM 音声信号波形を見る上で注意すべき点として，真ん中の三つは S/N 比が非常に悪く見える．これは皮膚に接着した聴診器型と異なりシリコーン型 NAM マイクロフォンは指で圧着する形式をとっていたためである. NAM マイクロフォンの表面自体がソフトであるため，指での固定の仕方や加える圧により微妙に感度や帯域が変化したため，試作品として最適固定法を模索中であった．しかしこの指による固定方式では図 3.12 の OCWHS 型，OCMSS 型，OCWSS 型などの無発声部に見られるように，感度が高ければ高いほど NAM マイクロフォン背部を押さえるために静止して力を入れた状態の指の軽微な振動（指の伸筋と屈筋のバランス）による接触性の雑音が入る．このノイズを除去するための固定方式は後述する．また 50Hz のハイパスフィルターを通すことで除去することも可能である.

皮膚接触感度が最も高いのは OCWSS タイプであった．OCMSS がそれに次ぐが, ソフトシリコーンを用いてもセラミック圧電素子を使った TMSS は極端に感度が低かった．ピエゾ素子も使用してみたが自作マイクアンプの 26dB 程度の増幅率では，NAM はサンプリングできなかった．ハードシリコーン型の OCWHS タイプと聴診器型の接触感度ほぼ同程度であった. OECM とソフトシリコーンの組み合わせが，高い接触感度をもたらすと言える.



皮膚接触面感度の順位	1	2	3	4	5
NAM マイクロフォンの型	OCWSS	OCMSS	聴診器型	OCWHS	TMSS

図 3.11　NAM マイクロフォンタイプ別皮膚接触面感度

3.6.3　外部雑音への頑強性（NMHF の気導音感度）

ソフトシリコーン型 NAM マイクロフォンを用いた，雑音に対する頑健性の数値的評価は，竹苗らの接話マイクと NAM マイクロフォンでの同時収録による比較実験[55]にて，60dBA，70dBA 程度の高レベル雑音下では，NAM マイクロフォンが接話マイクロフォンの認識性能を上回ることを示した．

マイクロフォン内部のコンデンサーマイクロフォンや圧電素子といったセンサーそのものの感度が高ければ，当然のことながら目的とする肉伝導 NAM も気導外部雑音もどちらもよく拾い，感度が低ければどちらもあまり入らない．外部雑音に対して頑強にするために肉伝導音マイクロフォンとして大切なことは，どれだけ皮膚接触面だけに感度を局在して集中させるかであり，それはマイクの構造と音媒体で決まる．

NAM マイクロフォンは肉伝導音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，若干ながら気導音外部雑音にも感度があるはずである．ただし感度の大きい接触面は使用時には人間の皮膚に密着している．そこで NAM マイクロフォン単体での皮膚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気導マイクロフォンとみなすことを提案する．これをここでは 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．つまりバーチャルな気導マイクロフォンとしての NMHF のインパルス応答，すなわちマイク特性を測定するのである．増幅率や出力レベル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の接触感度に応じてマイクアンプの増幅率や出力レベルを調節する．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気導

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．

環境の外部雑音が混入する道筋は二つある．ひとつは人体に接してないマイクの背部や側面にも気導音感度があり，ここから気導音としての外部雑音が混入する場合．もうひとつは体内伝導した外部雑音であり，これは NAM や BTOS と一緒に NAM マイクロフォンの感音面から入るものである．この NMHF の気導音感度という概念を用いれば, その二つの道筋のどちらからどれくらいの比率で外部雑音が混入しているのかは不明だが，その NAM マイクロフォンを人体に装着して実際に使用するときに，総合的にどのくらいの外部雑音が混入するかの目安となる．

また NAM マイクロフォンを皮膚に装着したときには，音声の生成系から直接振動をピックアップしているという理由で，同じ増幅率であれば音声を気導音声より大きく収録できるという利点（つまり増幅率や出力レベルを低く抑えられるという利点）と，外部雑音が人体というフィルタを一度通過するという利点の両方を総合的に評価できることになり，実用上便利である．

図 3.13 がこの評価法の概念図（上段）と実際のサンプル収録データである（下段）．気導音収録のための通常コンデンサマイクロフォンを耳元に置き，NAM マイクロフォンを最適装着位置に図 3.13 のように装着する．被験者の側方（マイク側）から約 50cm 離れた位置のスピーカーより規則的に一定間隔で TSP 信号を繰り返し流す. この際マイクアンプは同じ規格のものを用い，通常コンデンサマイクロフォン側は気導通常音声が適正にサンプリングされる出力レベルに，NAM マイクロフォン側は NAM 音が認識や聴取に適正な音量でサンプリングされるように出力レベルを調整する（0〜1MΩ）．

下段の二つの図が, 得られた音声信号波形と，FFT を使って得られる TSP 信号に対する周波数応答の図である．左側は気導通常音声を気導通常マイクロフォンで収録したものである．通常のマイクロフォン特性を測定する時と同じ方法であり，これが対照となる．右側は，NAM 音をある種類のソフトシリコーン型 NAM マイクロフォンで収録したものである．信号部分の FFT

は，NAM マイクロフォンを人体の頭部に装着した状態を一種の仮想気導マイクロフォンとみなして，その特性を測定しているわけであり，つまりNMHF の周波数応答の図である．NAM マイクロフォンの外部雑音に対する頑強性は，図の①の比較で TSP 信号波形の丈と幅が小さければ小さいほど，また図の②の比較でレスポンスが悪ければ悪いほど大きいことになる．

NAM マイクロフォンには，前述のごとく接触感音面から NAM や BTOS と同時に入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あるが，下段の二つの図の比較から分かるように，NAM 音を NAM マイクロフォンでサンプリングしたときの外部雑音は，通常マイクロフォンで通常音声をサンプリングするよりも低減する．

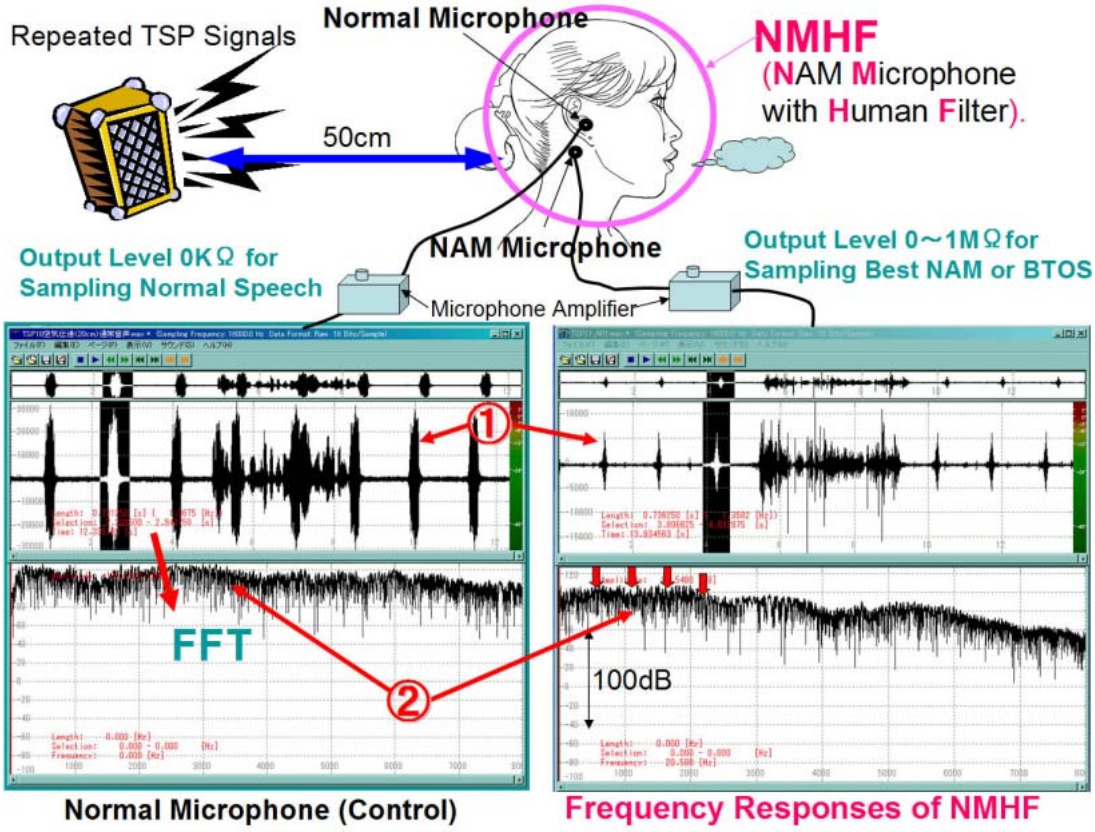


図 3.13　NMHF の概念と外部雑音への頑強性（NMHF の気導音感度）

62 ページ以降の図 3.15～3.19 は聴診器型, ハードシリコーン型(OCWHS 型), ソフトシリコーン型三種(OCMSS 型, OCWSS 型, TMSS 型)の NMHF の周波数応答の図である．NAM の適正音量に合わせた場合と BTOS の適正音量に合わせた場合を縦に並べて掲げた.

なお各図を見るときには対照として図 3.13 の左下の通常気導音コンデンサマイクロフォン収録の気導音 TSP による周波数応答の図と比較しながら見れば，その外部雑音耐性がよくわかるので，図 3.14 に拡大して再掲する.

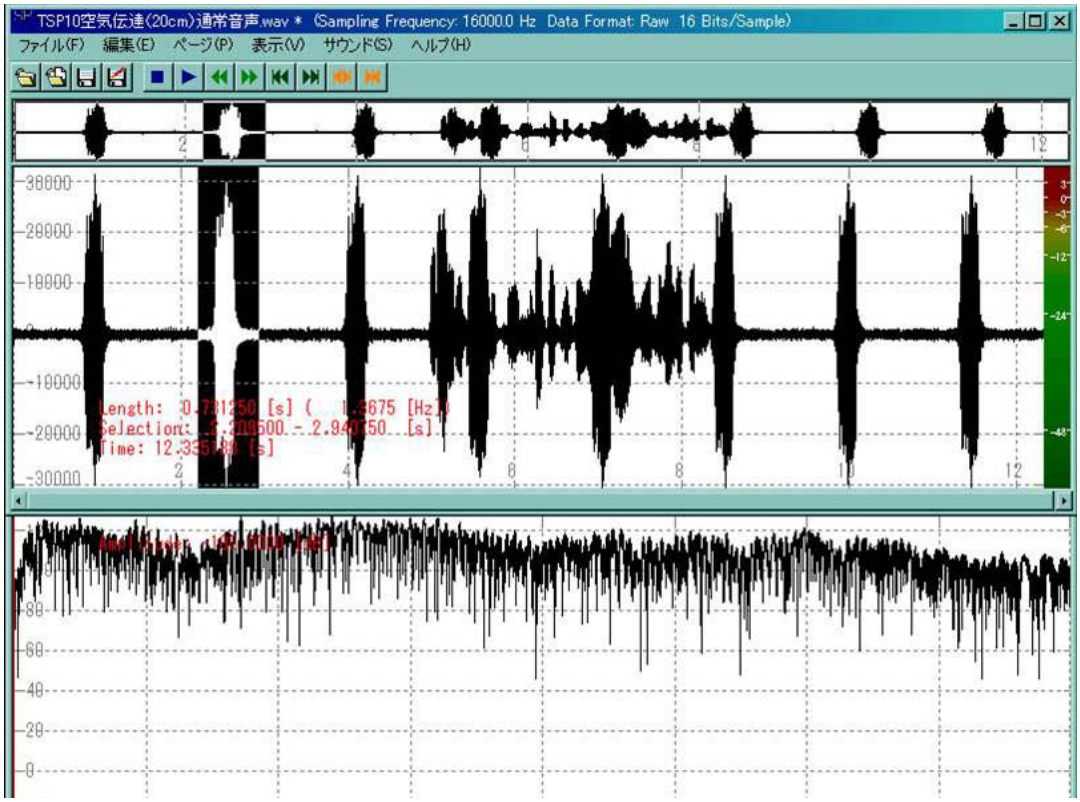


図 3.14　耳元のコンデンサマイクの気導音 TSP による周波数応答
（音声波形は気導通常音声）

図 3.15 より聴診器型では NAM 収録時, TSP 信号の振幅が約 40000. BTOS 収録時には 3000 程度となる．NMHF の周波数応答からは，音声認識に重要な 1KHz 周辺の帯域のノイズに脆弱であることがわかる.

図 3.16 から, ハードシリコーン型の接触感度は前述のように聴診器型とほぼ同等でありながら，NAM 収録時に TSP 信号振幅は約 12000，BTOS で約 2000 と聴診器型に比し外部雑音に頑強となる. マイクロフォンの構造上コンデンサマイクロフォンの周囲から空気を完全に排除した効果と考えられる.

図 3.17 の OCMSS 型は NAM で TSP 信号振幅は約 38000，BTOS では最大振幅は聴診器型より小さいが, 持続時間が長く平均的に雑音耐性は NAM, BTOS 収録時ともに聴診器型とあまり大差のない結果となっている.

図 3.18 でわかるように, OCWSS 型はソフトシリコーン NAM マイクロフォンの中で最も外部雑音に対して頑強な傾向にある．NAM で TSP 信号最大振幅は約 16000，BTOS では計測不能で TSP 信号をよく聞き取れない.

図 3.19 のセラミック圧電素子を使用した TMSS は図 3.9 と図 3.10 に示した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ったが, しかし接触面感度は OECM を用いたものに比べて低すぎ，図 3.19 でも，BTOS はともかく，使用したマイクアンプでは NAM 音は必ずしも十分な音量でサンプリングされない．適音量で NAM をサンプリングするには増幅率を大きく上げねばならず, NAM 音信号の振幅と TSP 信号の振幅を比較すれば，外部雑音耐性は OCWSS に比べて劣ることがわかり，聴診器型や OCMSS と同程度かそれ以下である.

一般に BTOS 信号を適音量でサンプリングする場合には, その NAM マイクロフォンを使ってもマイクアンプの増幅率や出力レベルを大幅に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 BTOS 音声信号の音量に比してかなり小さくなり，周波数応答も NAM 適正音量収録時より全体的に 20dB 程度低下させることができる.

結果的に雑音耐性に優れるのはハードシリコーン型（OCWHS 型）とソフトシリコーン型の OCWSS 型である.

外部雑音耐性の順位	1	2	3	4	5
NAM マイクロフォンの型	OCWSS	OCWHS	OCMSS	聴診器型	TMSS

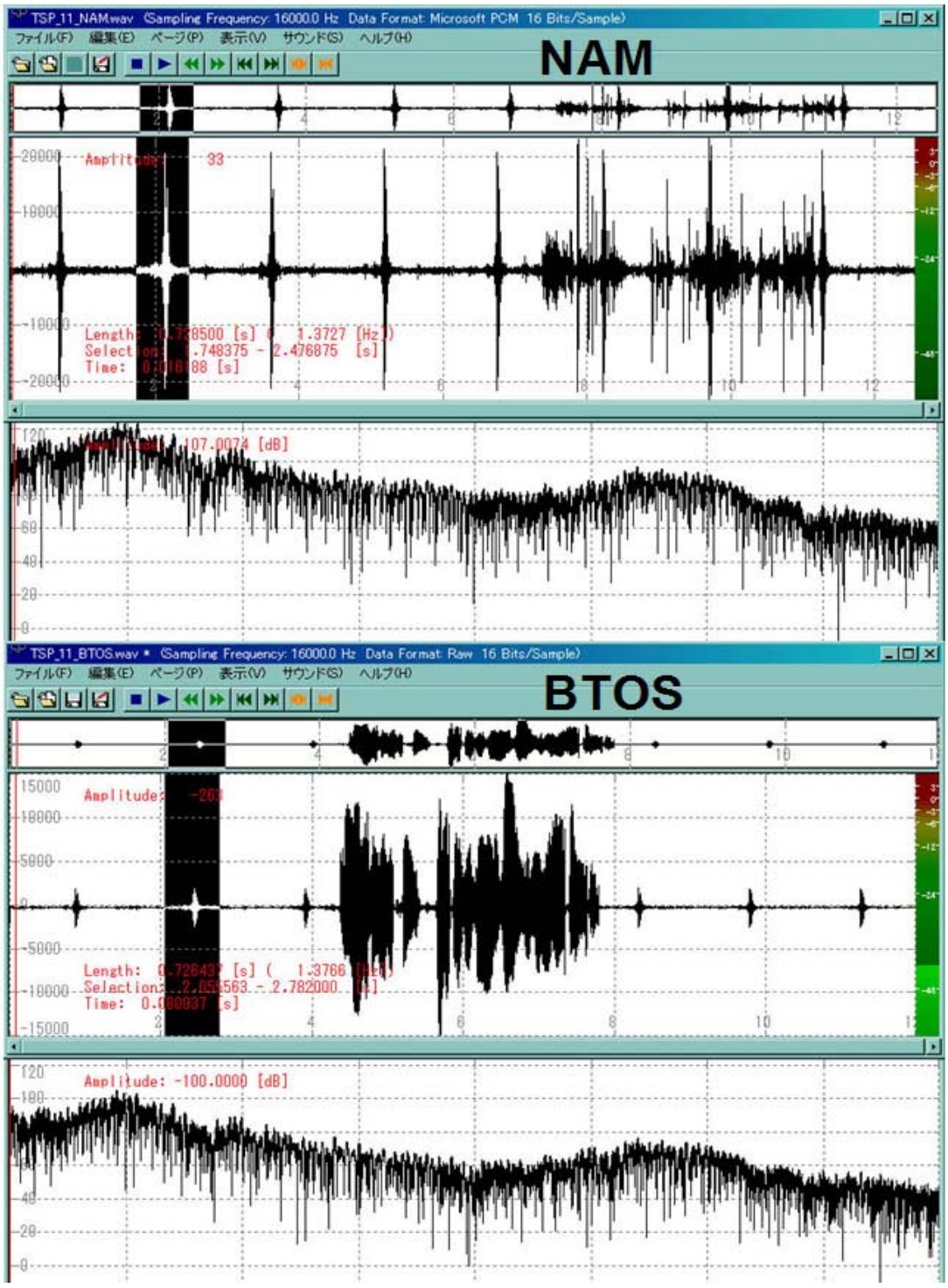


図 3.15　聴診器型 NAM マイクロフォンの NMHF 気導音感度

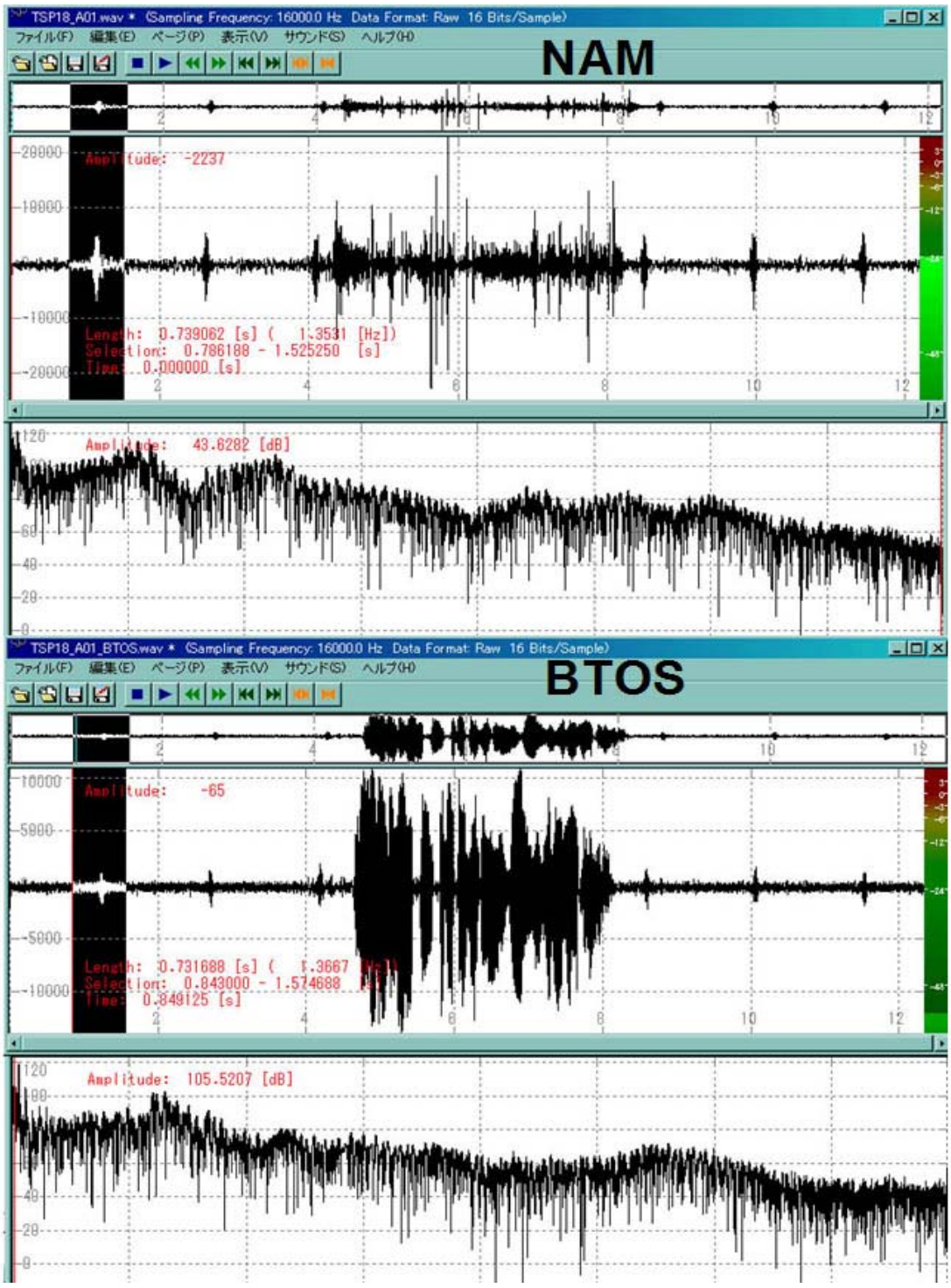


図 3.16　OCWHS 型の NMHF 気導音感度

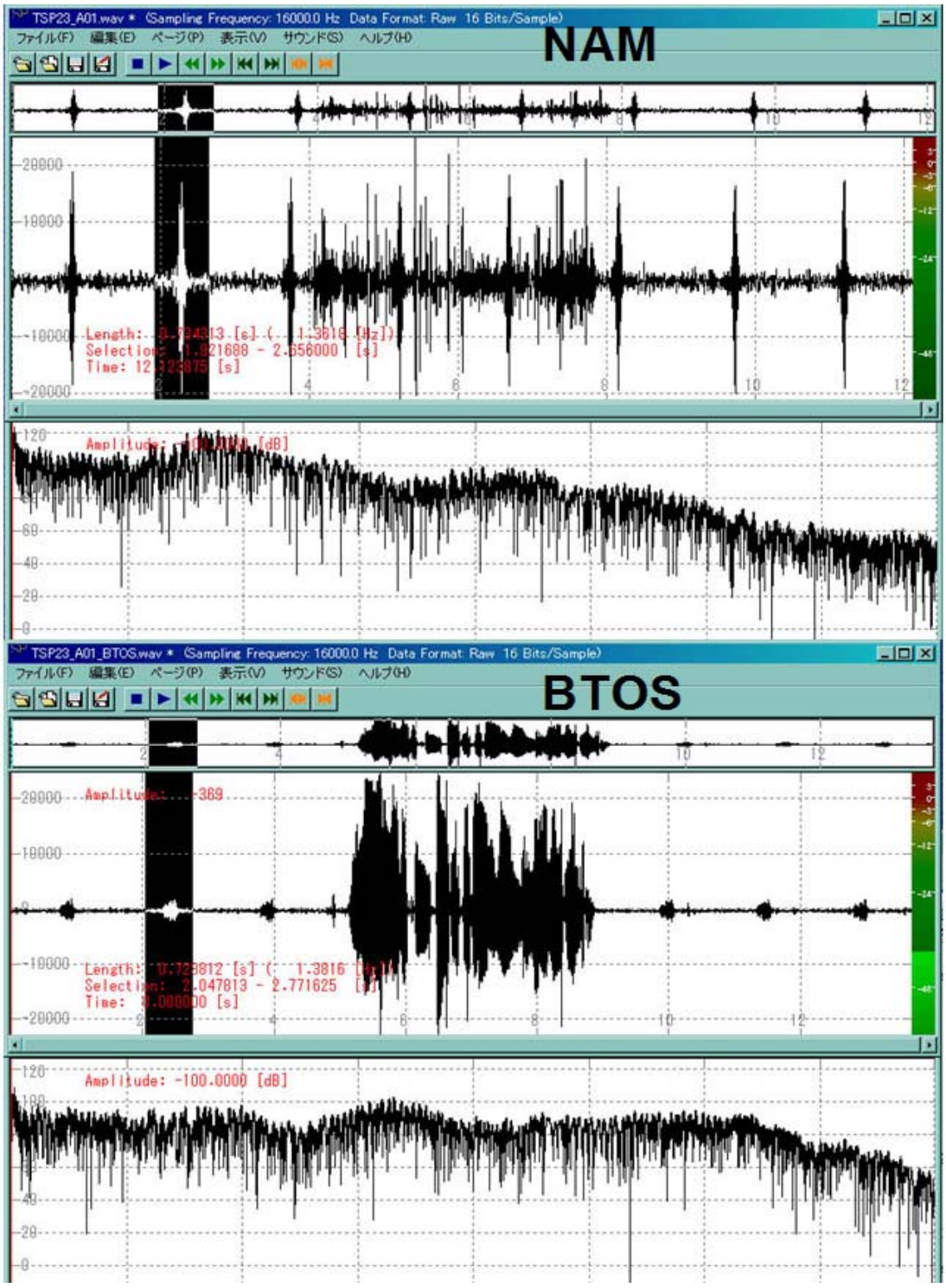


図 3.17　OCMSS 型の NMHF 気導音感度

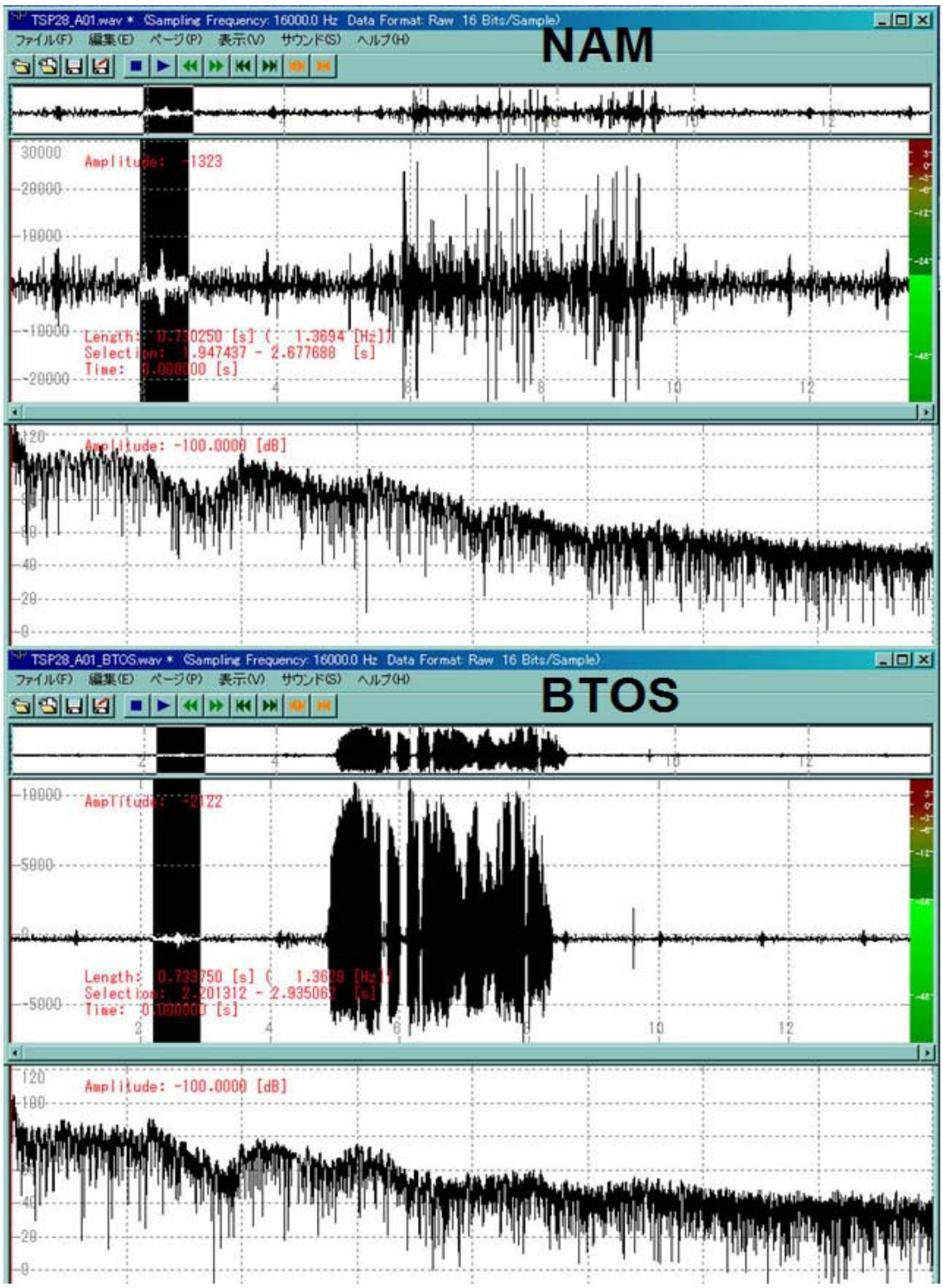


図 3.18　OCWSS 型の NMHF 気導音感度

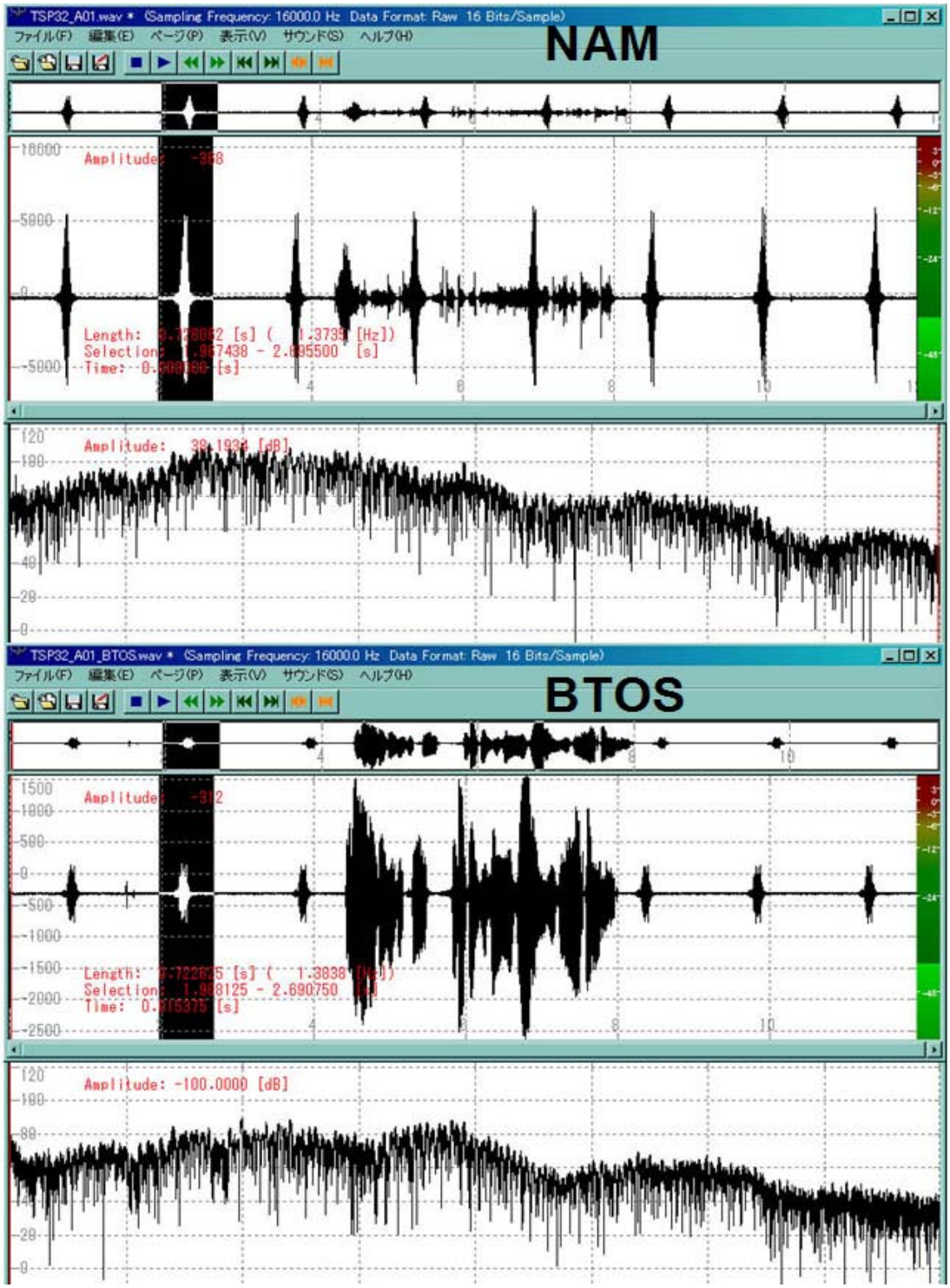


図 3.19　TMSS 型の NMHF 気導音感度

3.6.4　視覚的簡易評価のまとめ

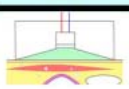
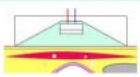
我々は現在までに数種類のシリコーンや樹脂素材，ゴム，レジンなどの素材とコンデンサマイクロフォンやセラミック圧電素子を用いて，多くの NAM マイクロフォン試作品を作成したが，その使用経験による特性からそれらを新，旧合わせて大きく 5 種類に分類した．その代表的なものをひとつずつ選んで, 各 NAM マイクロフォンを評価し, その大まかな傾向を述べた．

この 5 グループに帯域，皮膚接触面感度，外部雑音耐性の 3 点から，日頃の使用感も加えていささか強引に順位付けを試みると

- 帯域では TMSS，OCMSS，OCWSS，OCWHS，聴診器型の順でソフトシリコーン型の 3 種が有力である．
- 皮膚接触面感度では，OCWSS，OCMSS，聴診器型，OCWHS，TMSS の順で，OECM にソフトシリコーンを使用した 2 種の高さが目立つ．
- 外部雑音耐性では，OCWSS，OCWHS，OCMSS，聴診器型，TMSS の順で，OECM を同一素材媒体で包埋したものが特に頑強であった．

以上の結果と使用経験をふまえて現在はソフトシリコーンを用いた OCWSS 型と OCMSS 型をその用途に応じて使い分けている. 表 3.1 に NAM マイクロフォンタイプ別に特性を整理して掲げる．

表 3.1　NAM マイクロフォンの特性のまとめ

	聴診器型	ハードシリコーン型	ソフトシリコーン型三種		
	ST	OCWHS	OCMSS	OCWSS	TMSS
Band Width	0-2KHz	0-4KHz	0-6KHz	0-5KHz	0-7KHz
Contact Sensitivity	middle	low	middle	high	low
Noise Robustness	low	high	middle	high	low
構造の概念図					
使用しているセンサー	ECM	OECM	OECM	OECM	Transducer
			現行NAMマイクロフォン		

接触型マイクロフォンは，歴史的には相当以前から雑音下での通常音声収録目的のものが何種類かあったようであるが，全くといっていいほど普及してない．突発的にある研究者個人が一時的に研究するものの継続しなかったり，後継者が出なかったりしたようであり，その技術蓄積は気導音マイクロフォンの豊かな技術的広がりとは比べるべくもない．

接触型マイクロフォンの特性を定量的に提示する方法論も測定機器も現在の所確立されていないため(コンセンサスを得られるほど普及してないため)，できるだけ NAM マイクロフォンの使用経験のない人にも直感的かつ視覚的にわかりやすくなるように，NAM マイクロフォン製作と NAM や BTOS 収録の現場の経験と知恵から生まれた簡易評価方法を提示しながら，ソフトシリコーン NAM マイクロフォンの聴診器型に対する優位性を評価した．

図 3.20　NAM マイクロフォン工房

3.7　認識率による NAM マイクロフォンの評価

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

ソフトシリコーンを音媒体に用いた新 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 マイクロフォンと，旧式の聴診器型 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と評価用の 24 文[21]を NAM 発話にて同じ NAM マイクロフォンで読み上げて 16KHz サンプリングし，50Hz のハイパスフィルター処理をかけた．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）に HTK[64] を用い，350 文章，128 クラスターで 10 回の繰り返し話者適応（Iterative MLLR）[24]を行った．認識エンジンは Julius3.4[19] を用い，言語モデルの辞書として 20K 辞書[4]を用いた．認識率の評価は JDTK[21] を用い，単語認識精度を計算した．図 3.21 に MLLR の回数と単語認識精度を聴診器型 NAM マイクロフォンと比較して提示する．単語認識精度は聴診器型に比し約 5%の上昇を見た．

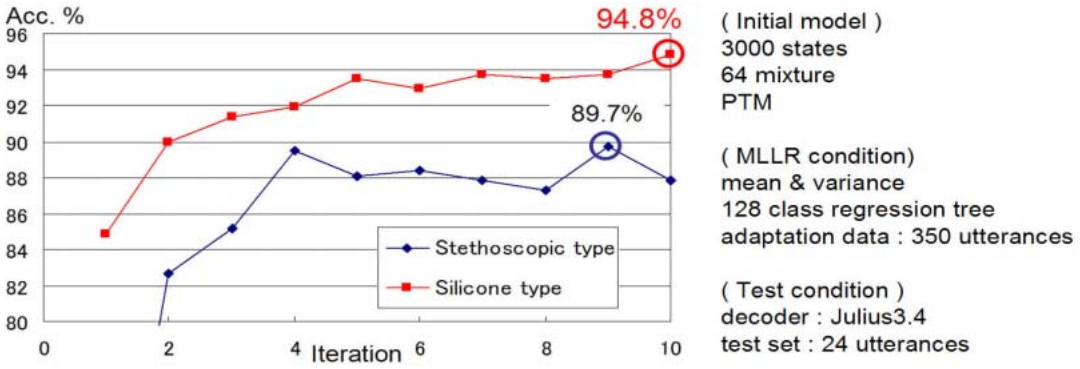


図 3.21　ソフトシリコーン型と聴診器型の NAM 認識率の比較

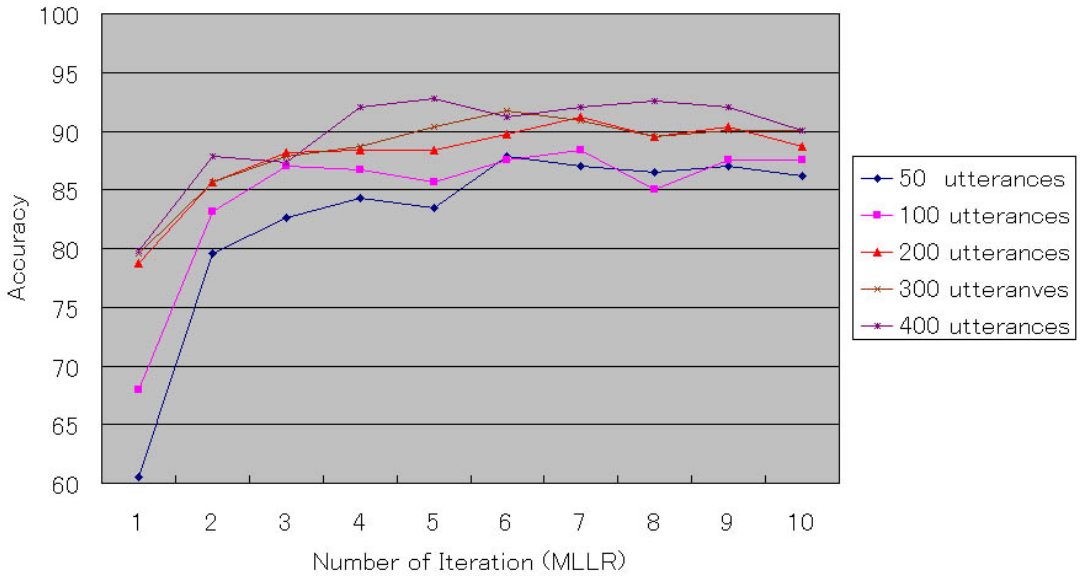


図 3.22　ソフトシリコーン型 NAM マイクロフォン（OCMSS）の Iterative MLLR における NAM 認識の適応文数の違いと認識率

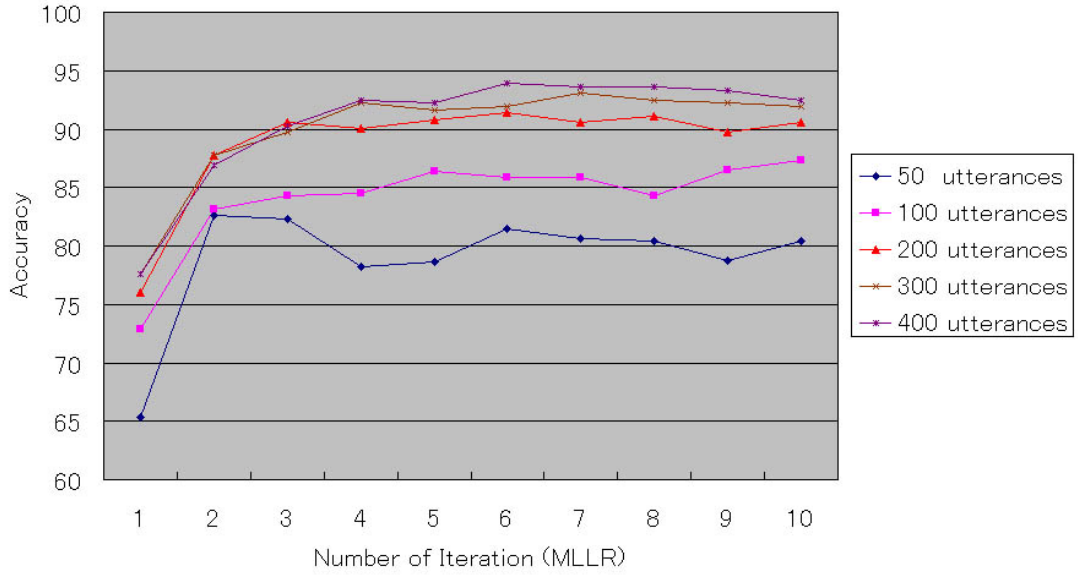


図 3.23　ソフトシリコーン型 NAM マイクロフォン（OCMSS）の Iterative MLLR における BTOS 認識の適応文数の違いと認識率

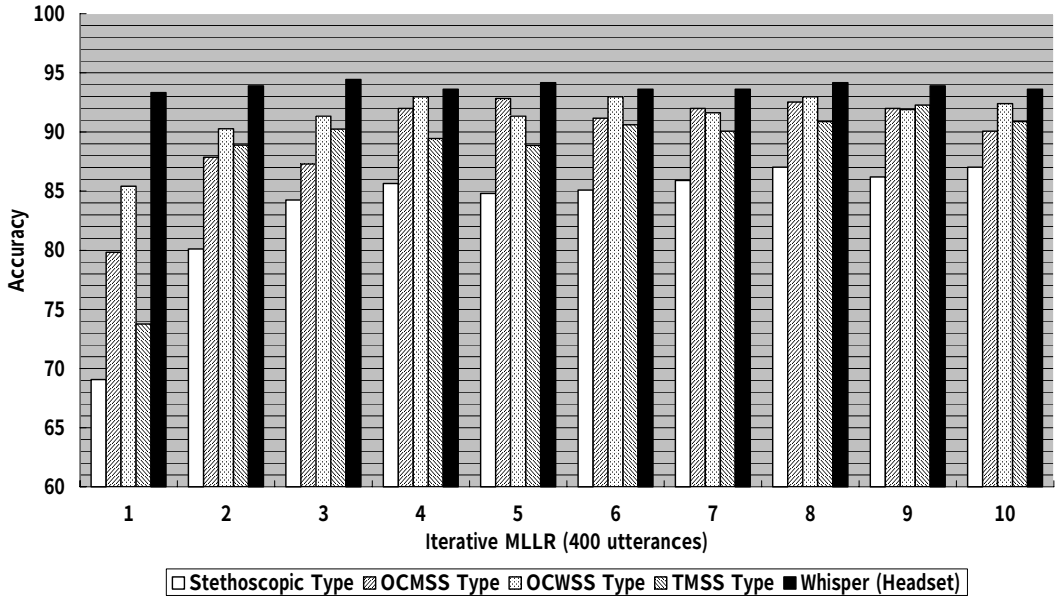


図 3.24　マイク別 NAM 認識率（Iterative MLLR 400 文章）
聴診器型，ソフトシリコーン型 3 種，気導音ささやき声（対照）の比較

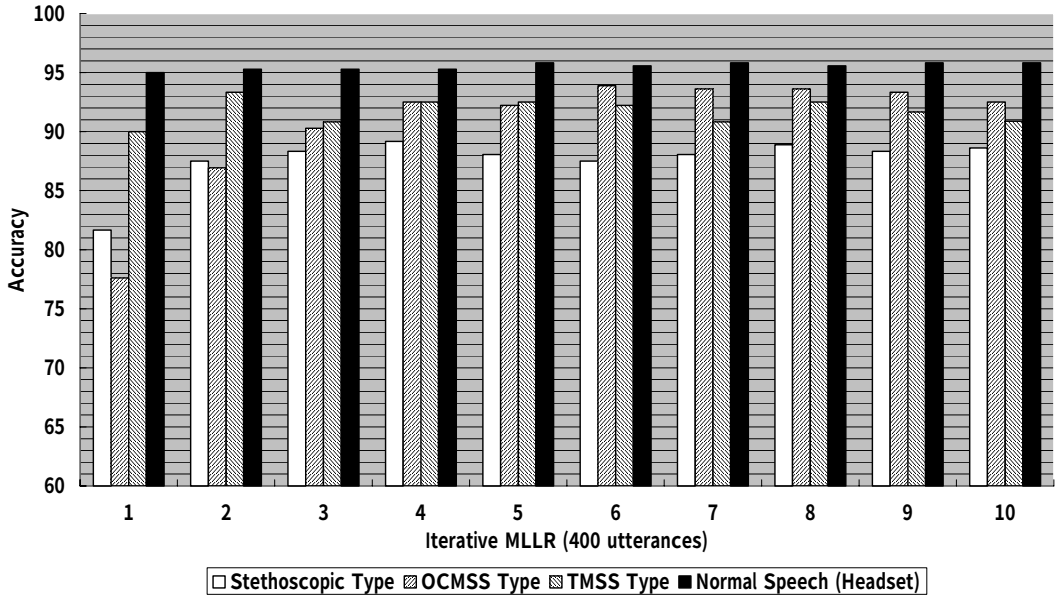


図 3.25　マイク別 BTOS 認識率（Iterative MLLR 400 文章）
聴診器型，ソフトシリコーン型 2 種，通常音声（対照）の比較

図 3.22 はソフトシリコーン型 NAM マイクロフォン（OCMSS 型）による NAM 認識で，話者適応（Iterative MLLR）に用いた読み上げ文数の違いによる単語認識精度を表すグラフである. 認識エンジン Julius のバージョンが 3.4.2 であること以外の条件は，図 3.20 の場合と全く同じである．図 3.23 は BTOS 認識の場合である．この二つ結果により，話者適応による NAM 音響モデル，BTOS 音響モデル作成には NAM，BTOS ともに聴診器型の場合とほぼ同様に 300 文～400 文の 6 回以上の繰り返し話者適応が妥当と考える.

図 3.24 はソフトシリコーン型 NAM マイクロフォンでサンプリングした NAM と聴診器型 NAM マイクロフォ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フォンで NAM とは別に収録したささやき声の話者適応(Iterative MLLR)による認識率の比較を示すグラフである．ここでも Julius3.4.2 を認識エンジンに用いていること以外の条件は同じである．繰り返し回数は 10 回，適応文数は 400 文章である．OCMSS 型，OCWSS 型，TMSS 型ともにソフトシリコーン型 NAM マイクロフォンは聴診器型 NAM マイクロフォンの単語認識精度を上回る．図 3.25 は対照にヘッドセットマイクロフォンで収録した通常音声を用いた，BTOS の話者適応による認識率比較を示す. ここでも OCMSS 型, TMSS 型のソフトシリコーン型 NAM マイクロフォンは聴診器型の単語認識精度を上回った．OCCWSS 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，このグラフには含めなかった.

図 3.24 でも図 3.25 でも対照（黒い棒グラフ）の気導音声と比べると，やはり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がわからない．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である．肉伝導音声の認識に適したパラメータ抽出法を検討することも，もうひとつの大きな課題である.

3.8　聴取実験による NAM マイクロフォンの評価

聴診器型 NAM マイクロフォンを用いてサンプリングした NAM も，初めて聞いてわかる場合もあるし，何度も聞いているうちに耳に慣れが生じて，かなり内容がわかるようになるものである．

新しいソフトシリコーン伝導型 NAM マイクロフォンでは，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言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，気導音声への接近を目標とする．そこでソフトシリコーン伝導型 NAM マイクロフォ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．

3.8.1　聞き取り実験の方法

まず聴き取ってもらう対象として，十数単語程度の文章（日常的な用件文が主体）を 12 個，意味のわかる単独の単語のみを 12 個，そしてまったく意味をなさない 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知らないような語彙は排除した．つまり一般人の言語モデルに必ず存在すると思われる言葉を選んだ．言語モデルに存在しない未知語の音声（音韻・音素）の認識は無意味単語で評価する．表 3.2 にその文章と単語を掲示する．

表 3.2　聞き取りテストの読み上げ文と単語

I. 意味文章（日常的な用件文）

１．会議が始まる前に，携帯電話の電源を切って下さい.
２．11 日の同窓会には，葉書とアルバムを持ってきてください.
３．明日のミーティングは，午後７時から第二会議室で行います.
４．お金は昨日の午前中に，指定の口座に振り込みました.
５．帰りにスーパーでチューブ入りのわさびを買ってきてね.
６．私の父は昭和１７年の生まれで，戦後の人間です.
７．書店に注文している本が，明日の夕方に届くはずです.
８．五千円からですと，おつりは 247 円になります.
９．予約したチケットの料金は，コンビニからでも払えるよ.
10. 頭が痛くて熱が出てきたので，今日は学校を休むことにする.
11. 犬は飼うなと叱られたので，鳥を飼うことにしました.
12. もう少し明るい場所で，きちんと座って本を読みなさい.

II. 意味単語（日常的なものからやや抽象的なものまで）

１．腕時計　２．カーテン　３．爪切り　４．弁護士　５．あじさい　６．双子座
７．模擬テスト　８．上流社会　９．図書館　10. 町並み　11. 街路樹　12. コーヒー

III. 無意味単語（ひらがなにして 3〜4 文字程度の無意味な単語）　例.「まりにょ」

１．かがら　２．むぶぷ　３．のとど　４．えほしゃ　５．うふに　６．ねっけ
７．かーちゅ　８．りょっき　９．さーじつ　10. ほびゅぱ　11. ちしじ　12. なだた

テストセットは以上の 36 問題を，それぞれの文や単語ごとに男性話者一名が標準的な速度で読み上げて，新旧 2 種類の NAM マイクロフォンで録音した．対照とするため肉伝導音の他に気導音の通常音声やささやき声もヘッドセットマイクロフォンにて録音した．すべてについて 8KH z と 16KHz の 2 種類のサンプリングレートでデジタル化した(8KHz は 16KHz から作成)．

表 3.3 に録音サンプルの種類をまとめる．なおソフトシリコーン伝導型 NAM マイクロフォンとしては，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．

音量は音声信号の最大振幅を見て，大きく違わない程度に配慮した．

表 3.3　録音サンプルの種類

略号	対象音声	マイクロフォン	サンプリングレート
NA_ST_08	NAM	聴診器型NAMマイクロフォン	8KHz
NA_ST_16	NAM	聴診器型NAMマイクロフォン	16KHz
BT_ST_08	BTOS	聴診器型NAMマイクロフォン	8KHz
BT_ST_16	BTOS	聴診器型NAMマイクロフォン	16KHz
NA_OC_08	NAM	ソフトシリコーン型NAMマイクロフォン	8KHz
NA_OC_16	NAM	ソフトシリコーン型NAMマイクロフォン	16KHz
BT_OC_08	BTOS	ソフトシリコーン型NAMマイクロフォン	8KHz
BT_OC_16	BTOS	ソフトシリコーン型NAMマイクロフォン	16KHz
WS_HS_08	ささやき声	ヘッドセットマイクロフォン	8KHz
WS_HS_16	ささやき声	ヘッドセットマイクロフォン	16KHz
NS_HS_08	通常音声	ヘッドセットマイクロフォン	8KHz
NS_HS_16	通常音声	ヘッドセットマイクロフォン	16KHz

無作為に様々な録音のサンプルを聴取してもらうため，被験者 12 名用に 12 グループのテストセットを作成した. それぞれのグループは同じ文章や単語を同じ順に聴取させるが, 同じ問題については表 3.3 に挙げた 12 種類の録音の並び方を無作為の順列に選ぶ．こうして作ったのが表 3.4 の録音サンプルの割当表である．表の各行について 12 種類の録音をすべて登場させて，しかも並び方を無作為にする．これで各人は 12 種類の録音をランダムに聴

くことになる．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．

表 3.4　問題に対する録音サンプル割当表（各行について順列はランダム）

問題番号	1	2	3	4	5	6	7	8	9	10	11	12	単語数
101	BT_ST_16	BT_OC_16	NA_OC_08	NS_HS_08	WS_HS_08	NA_ST_16	WS_HS_16	NA_ST_08	NS_HS_16	NA_OC_16	BT_OC_08	BT_ST_08	12
102	NS_HS_16	NS_HS_08	WS_HS_08	NA_OC_16	NA_ST_16	BT_OC_16	BT_ST_16	NA_OC_08	BT_ST_08	WS_HS_16	BT_OC_08	NA_ST_08	13
103	WS_HS_08	NA_OC_16	NA_OC_08	NS_HS_08	BT_ST_08	BT_OC_16	NA_ST_16	NA_ST_08	NS_HS_16	BT_ST_16	WS_HS_16	BT_OC_08	13
104	NA_ST_08	WS_HS_08	NS_HS_08	BT_ST_16	NA_OC_16	NS_HS_16	BT_OC_16	NA_OC_08	NA_ST_16	BT_ST_08	BT_OC_08	WS_HS_16	12
105	WS_HS_08	BT_ST_08	NA_ST_16	NS_HS_16	NA_OC_08	BT_OC_16	NS_HS_08	NA_OC_16	NA_ST_08	WS_HS_16	BT_ST_16	BT_OC_08	13
106	NA_OC_08	BT_OC_08	NS_HS_16	BT_ST_08	NA_ST_16	WS_HS_16	WS_HS_08	NA_OC_16	BT_OC_16	NS_HS_08	BT_ST_16	NA_ST_08	14
107	NA_ST_16	BT_OC_08	NS_HS_08	NA_OC_16	WS_HS_08	BT_OC_16	WS_HS_16	BT_ST_08	NA_ST_08	NS_HS_16	NA_OC_08	BT_ST_16	13
108	NA_ST_08	NA_OC_08	WS_HS_08	NS_HS_16	WS_HS_16	BT_ST_08	BT_ST_16	BT_OC_16	NS_HS_08	NA_ST_16	NA_OC_16	BT_OC_08	11
109	NA_OC_16	WS_HS_08	BT_OC_08	NS_HS_16	NA_OC_08	BT_ST_16	NA_ST_16	NA_ST_08	BT_ST_08	NS_HS_08	BT_OC_16	WS_HS_16	11
110	BT_OC_08	NA_OC_08	NA_OC_16	BT_OC_16	NA_ST_16	BT_ST_16	NS_HS_16	NS_HS_08	BT_ST_08	NA_ST_08	WS_HS_16	WS_HS_08	17
111	BT_OC_08	NS_HS_08	BT_ST_08	NA_OC_16	NA_ST_16	NS_HS_16	WS_HS_08	NA_ST_08	NA_OC_08	WS_HS_16	BT_ST_16	BT_OC_16	16
112	WS_HS_08	NA_OC_16	BT_ST_16	NS_HS_16	BT_OC_16	NA_OC_08	NS_HS_08	WS_HS_16	NA_ST_16	NA_ST_08	BT_ST_08	BT_OC_08	11
201	NA_OC_08	NA_OC_16	BT_ST_16	WS_HS_16	NA_ST_16	BT_OC_16	NS_HS_16	BT_OC_08	BT_ST_08	NS_HS_08	WS_HS_08	NA_ST_08	1
202	WS_HS_08	WS_HS_16	NS_HS_16	BT_ST_16	BT_OC_16	NS_HS_08	BT_ST_08	BT_OC_08	NA_ST_08	NA_OC_16	NA_ST_16	NA_OC_08	1
203	NA_OC_08	BT_ST_16	WS_HS_08	NA_ST_16	BT_OC_16	NA_OC_16	NS_HS_16	WS_HS_16	BT_OC_08	NS_HS_08	BT_ST_08	NA_ST_08	1
204	NS_HS_08	BT_ST_08	NA_ST_08	BT_OC_16	NS_HS_16	WS_HS_08	NA_ST_16	BT_OC_08	WS_HS_16	BT_ST_16	NA_OC_16	NA_OC_08	1
205	NA_ST_16	WS_HS_16	WS_HS_08	NS_HS_08	BT_OC_16	BT_ST_08	NA_ST_08	BT_ST_16	BT_OC_08	NS_HS_16	NA_OC_08	NA_OC_16	1
206	NS_HS_16	NS_HS_08	NA_OC_08	WS_HS_08	WS_HS_16	BT_OC_08	BT_OC_16	BT_ST_16	NA_ST_08	BT_ST_08	NA_ST_16	NA_OC_16	1
207	NS_HS_08	NS_HS_16	NA_ST_16	WS_HS_16	BT_OC_16	BT_OC_08	NA_ST_08	NA_OC_08	BT_ST_16	NA_OC_16	WS_HS_08	BT_ST_08	1
208	WS_HS_08	NA_ST_16	NA_ST_08	BT_ST_16	BT_ST_08	NA_OC_08	NS_HS_08	NA_OC_16	BT_OC_08	NS_HS_16	WS_HS_16	BT_OC_16	1
209	BT_OC_08	NS_HS_08	NA_OC_16	NA_ST_16	BT_ST_16	WS_HS_08	BT_ST_08	BT_OC_16	WS_HS_16	NS_HS_16	NA_OC_08	NA_ST_08	1
210	NA_OC_16	NA_ST_08	BT_OC_08	NS_HS_16	BT_ST_16	NA_OC_08	BT_ST_08	WS_HS_16	WS_HS_08	BT_OC_16	NA_ST_16	NS_HS_08	1
211	BT_ST_08	WS_HS_16	BT_OC_08	WS_HS_08	NS_HS_16	NS_HS_08	NA_OC_08	BT_ST_16	BT_OC_16	NA_OC_16	NA_ST_16	NA_ST_08	1
212	BT_ST_16	NA_ST_08	WS_HS_08	NS_HS_08	NA_OC_08	NA_OC_16	NA_ST_16	BT_ST_08	BT_OC_08	NS_HS_16	BT_OC_16	WS_HS_16	1
301	BT_OC_08	WS_HS_16	BT_ST_08	BT_OC_16	NS_HS_08	NA_OC_08	BT_ST_16	NA_OC_16	NA_ST_16	NA_ST_08	NS_HS_16	WS_HS_08	1
302	BT_OC_08	BT_ST_16	NA_ST_08	NS_HS_08	NA_OC_16	WS_HS_08	NS_HS_16	WS_HS_16	BT_OC_16	NA_OC_08	BT_ST_08	NA_ST_16	1
303	WS_HS_16	NA_ST_16	BT_OC_08	NA_OC_16	NS_HS_08	NA_OC_08	WS_HS_08	BT_ST_16	BT_OC_16	NA_ST_08	NS_HS_16	BT_ST_08	1
304	BT_OC_08	BT_OC_16	WS_HS_16	WS_HS_08	BT_ST_16	NS_HS_08	NA_ST_08	BT_ST_08	NS_HS_16	NA_OC_08	NA_ST_16	NA_OC_16	1
305	BT_ST_16	NS_HS_08	BT_OC_08	BT_ST_08	NS_HS_16	NA_ST_16	WS_HS_08	WS_HS_16	BT_OC_16	NA_OC_16	NA_ST_08	NA_OC_08	1
306	BT_ST_16	WS_HS_08	NA_ST_08	BT_OC_08	BT_ST_08	NS_HS_16	NA_OC_08	NS_HS_08	WS_HS_16	NA_OC_16	BT_OC_16	NA_ST_16	1
307	NA_OC_08	WS_HS_16	BT_ST_16	BT_OC_08	NA_OC_16	NS_HS_08	NA_ST_16	NS_HS_16	BT_OC_16	BT_ST_08	NA_ST_08	WS_HS_08	1
308	NA_OC_16	BT_ST_16	BT_OC_08	NA_OC_08	NS_HS_16	WS_HS_08	NA_ST_08	NA_ST_16	BT_ST_08	BT_OC_16	NS_HS_08	WS_HS_16	1
309	BT_OC_16	NA_ST_16	WS_HS_16	WS_HS_08	NS_HS_16	BT_OC_08	NA_OC_16	NS_HS_08	BT_ST_16	NA_ST_08	BT_ST_08	NA_OC_08	1
310	WS_HS_16	NS_HS_16	WS_HS_08	NA_OC_08	BT_OC_16	NA_OC_16	BT_OC_08	NA_ST_08	BT_ST_16	NS_HS_08	BT_ST_08	NA_ST_16	1
311	BT_ST_16	WS_HS_08	BT_OC_16	NS_HS_08	NA_OC_16	BT_ST_08	BT_OC_08	WS_HS_16	NS_HS_16	NA_ST_08	NA_OC_08	NA_ST_16	1
312	BT_ST_08	NS_HS_08	BT_OC_08	NA_OC_16	NS_HS_16	BT_ST_16	WS_HS_08	NA_ST_08	NA_OC_08	WS_HS_16	NA_ST_16	BT_OC_16	1
年齢性別	M_45	M_59	M_24	M_15	M_39	F_69	F_32	F_24	F_45	F_13	M_69	F_50	

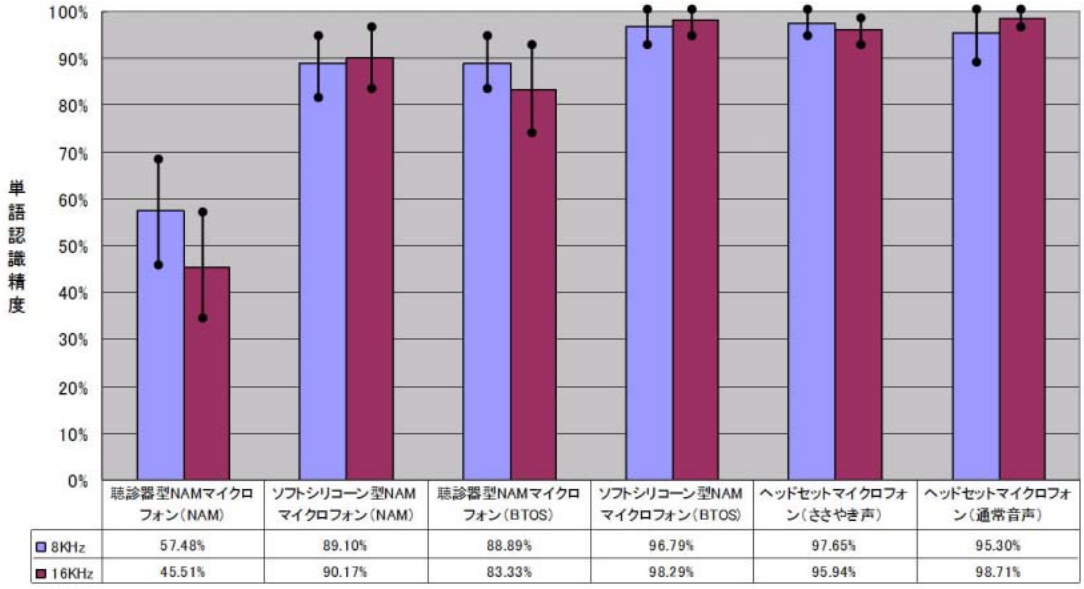
被験者として NAM や BTOS を聞き慣れていない 10 代から 60 代までの男性 6 名，女性 6 名の計 12 名を選出した．世代ごとに男女各一名ずつを原則に選び平均年齢は男 41.83 歳, 女 38.83 歳であった．これをまたクジでランダムに 12 のテストセットに割り振った．

テストの形式は，ひとつの問題について同じ録音を全部三回ずつ密閉式ヘッドフォンで聴取してもらい，聴いた結果をその度ごとに答案用紙に書き取ってもらうこととした．全員が 12×3×3 の 108 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．N：全単語数，D：脱落誤り数，S：置換誤り数，I：挿入誤り率として単語認識精度は$\frac{N-D-S-I}{N}$として計算する．

3.8.2　実験結果

まず各発声法と各マイクロフォ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に有意な差は見られず，聴診器型 NAM マイクロフォンにおいてはむしろ 16KHz で認識精度が下がる傾向が見られたのは興味深かった．いずれにせよ 8KHz サンプリングの携帯電話音質でも受話者の聞き取り精度に影響はあまりないことがわかる（図 3.26）．

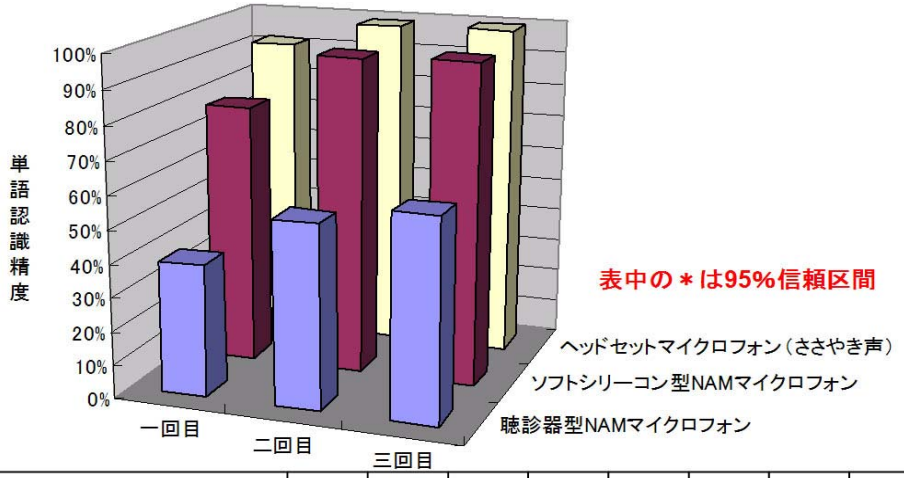


	聴診器型NAMマイクロフォン（NAM）	ソフトシリコーン型NAMマイクロフォン（NAM）	聴診器型NAMマイクロフォン（BTOS）	ソフトシリコーン型NAMマイクロフォン（BTOS）	ヘッドセットマイクロフォン（ささやき声）	ヘッドセットマイクロフォン（通常音声）
8KHz	57.48%	89.10%	88.89%	96.79%	97.65%	95.30%
16KHz	45.51%	90.17%	83.33%	98.29%	95.94%	98.71%

図 3.26　各収録法のサンプリングレート別の聞き取り認識率
（縦線は 95%信頼区間）

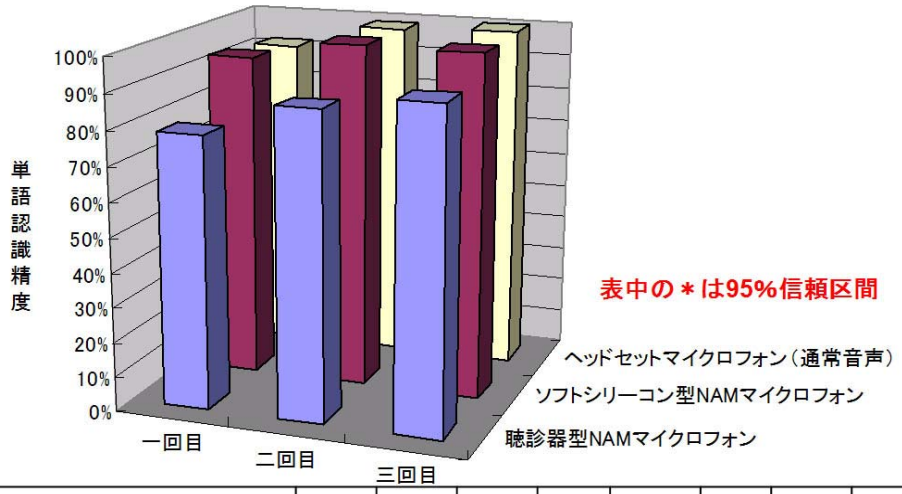
各収録方法でサンプリングレート間の有意な差が見られないことをふまえて，サンプル数を増やして認識精度の統計的信頼性を増すため，両者を混合して，文章の単語認識精度を計算した．対照としてヘッドセットマイクロフォンで収録したささやき声（明らかに気導音としての音量は気導 NAM 発声音より大きい）を対比して図 3.27 に掲げる．

同じサンプルを被験者は 3 回繰り返して聴くので, 当然一回目より二回目, 二回目よりも三回目が認識精度は上昇する．NAM の聞き取りについては旧式の聴診器型 NAM マイクロフォンに比して, ソフトシリコーン伝導型 NAM マイクロフォ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二回目，三回目の聞き取りについては有意差がなかった．一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も原因にあると思われる．二回目以降は一回目の記述をふ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．



NAM	一回目	＊	二回目	＊	三回目	＊	総計	＊
聴診器型NAMマイクロフォン	39.42%	14.82%	54.81%	15.75%	60.26%	14.09%	50.68%	8.45%
ソフトシリコーン型NAMマイクロフォン	78.20%	11.39%	94.87%	5.82%	95.83%	5.41%	88.86%	4.83%
ヘッドセットマイクロフォン(ささやき声)	91.67%	5.91%	99.04%	1.51%	99.68%	0.78%	96.68%	2.14%

図 3.27　NAM による文章聞き取りの認識率（対照：ささやき声）



BTOS	一回目	＊	二回目	＊	三回目	＊	総計	＊
聴診器型NAMマイクロフォン	78.52%	12.24%	87.82%	9.40%	91.99%	6.17%	85.58%	5.44%
ソフトシリコーン型NAMマイクロフォン	93.91%	5.19%	99.36%	1.00%	99.36%	1.00%	97.35%	1.82%
ヘッドセットマイクロフォン（通常音声）	91.99%	8.32%	99.03%	1.00%	100.00%	0.00%	96.85%	2.82%

図 3.28　BTOS による文章聞き取りの認識率（対照：通常音声）

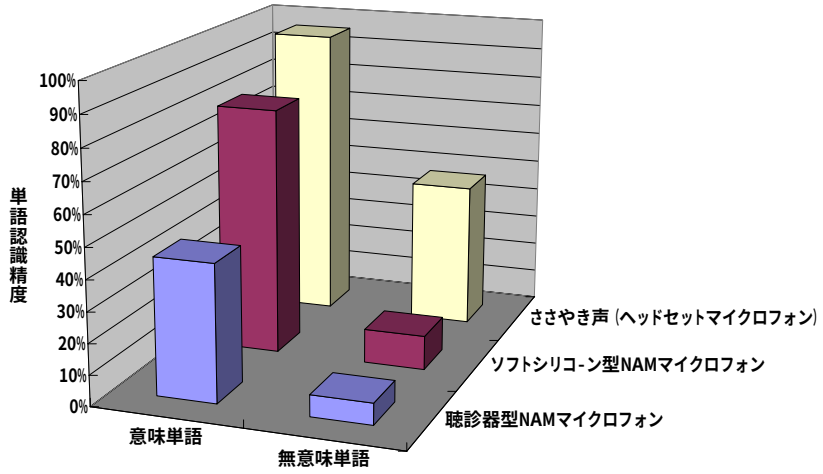
図 3.28 に見られるように, 有声子音, 無声子音の鑑別が容易であるためか, 各収録法で BTOS の方が NAM より高い単語認識精度を示す. BTOS でもソフトシリコーン伝導型 NAM マイクロフォンは，対照であるヘッドセットマイクロフォン収録通常音声の単語認識精度と有意差はなく，聴診器型 NAM マイクロフォンに比べて二回目，三回目の聞き取りで有意に高かった.

以上が 12 文章，12 人，三回聞き取りの結果であるが，文章の場合，機械認識と異なり，人間は言語モデルに匹敵する語彙を持つ以外にも，文脈から文章を類推することが可能である.

そこで単独単語認識であるが，これは言語モデルこそ使えるが，文脈がなく，より高次の知能でこれを補うことができない．したがって単独単語の聞き取りの方が, より機械認識のパターンに近い認識の仕方を行うことになる.

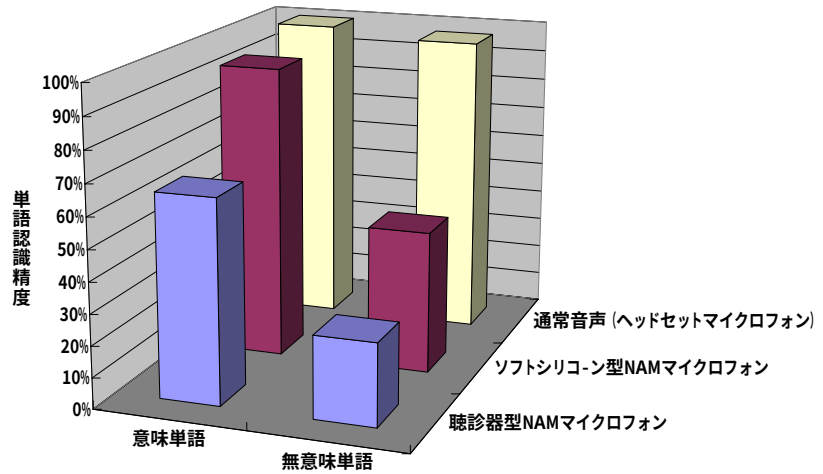
また無意味単語では，言語モデルは通用せず，まったく音韻，音素の聞き取りとなる．ひとつでも音素を聞き違えると，置換誤りとして単語認識精度を計算した．そのためこのテストが最も厳しい条件のテストとなっている.

図 3.29 と図 3.30 に NAM と BTOS の単独単語の単語認識精度をしめす.



	意味単語	無意味単語
聴診器型NAMマイクロフォン	44.44%	6.94%
ソフトシリコ-ン型NAMマイクロフォン	80.56%	11.11%
ささやき声（ヘッドセットマイクロフォン）	95.83%	47.22%

図 3.29　NAM の単独単語の単語認識精度



	意味単語	無意味単語
聴診器型NAMマイクロフォン	65.28%	26.39%
ソフトシリコ-ン型NAMマイクロフォン	94.44%	45.83%
通常音声（ヘッドセットマイクロフォン）	100%	97.22%

図 3.30　BTOS の単独単語の単語認識精度

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.

NAM でも BTOS でも，文章や意味単語においては気導音声にかなり接近したソフトシリコーン型 NAM マイクロフォンも，無味単語になると極端にその認識率を落とす. 無意味単語ではもともと気導音声のささやき声でさえ，認識率が 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通常音声とささやき声にこれほど大きな認識率の違いがある．ただこの低さは問題に意図的に同一単語内で有声子音，無声子音の判別を要する問題を作成したことに起因しているかもしれない.

3.8.3 聞き取り実験のまとめ

NAM マイクロフォンでは，BTOS においても NAM においても，旧式の聴診器型よりもソフトシリコーン伝導型 NAM マイクロフォンが，人間の聞き取りにおいて認識の精度を大きく高めることがわかった，また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ような状況においては，認識率は特に NAM において気導音声にまだまだかなり劣ることがわかった. 携帯電話などでの会話中にも，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，NAM マイクロフォンのさらなる改良によって是非ともこの音素の判別性を上げなければならない. また音素までの判別，つまり無意味単語の単語認識率を上げることが NAM の通常音声化の研究，BTOS の信号処理による音質変化の研究の重要な課題であると考える.

そして気導音と肉導音の，人間の音素聞き取りに対する情報量の差異は，スペクトラムで違いが目立つ高域というよりもむしろ，サンプリングレート 8KHz の携帯電話帯域である 4KHz まで帯域のどこかに隠されている可能性もある. 参考のために 8KHz サンプリングした NAM，BTOS，ささやき声，

通常音声を以下の図 3.31, 図 3.32 に掲げる. 発話内容は問題番号 I－1 の「会議が始まる前に，携帯電話の電源を切ってください」である．いずれも上段が肉伝導音声の NAM と BTOS の音質向上である.

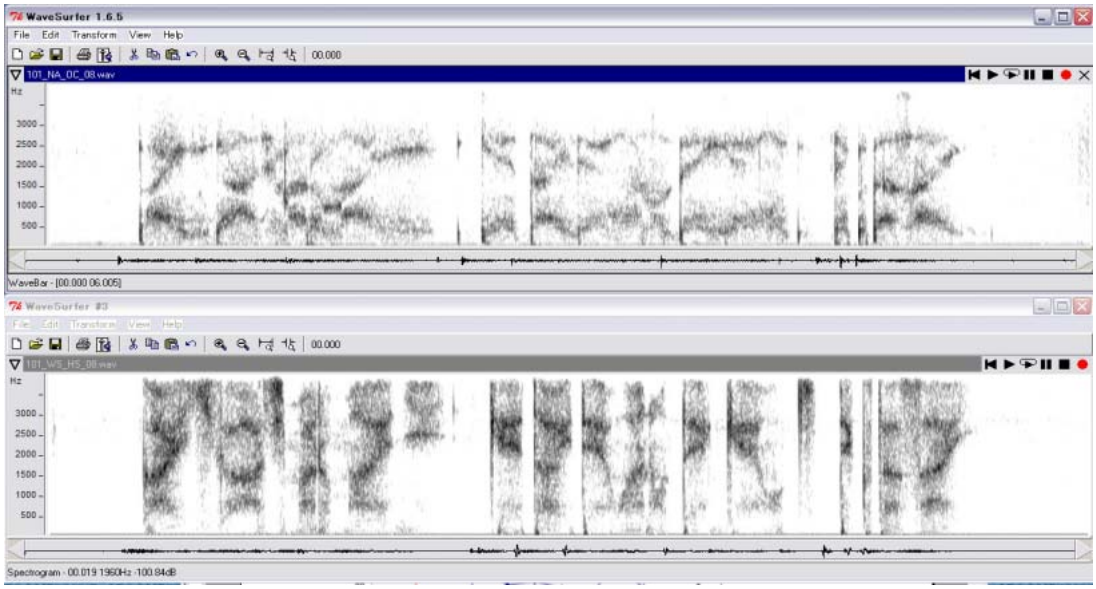


図 3.30　8KHz サンプリング NAM とささやき声のスペクトラム

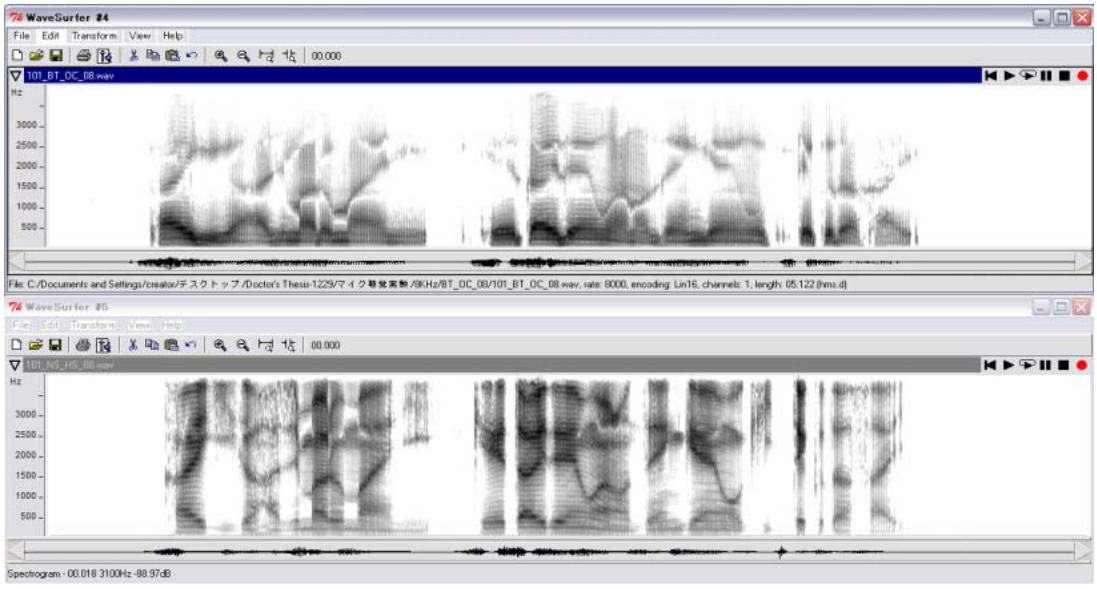


図 3.32　8KHz サンプリング BTOS と通常音声のスペクトラム

3.9　新 NAM マイクロフォンの工夫

3.9.1　NAM マイクロフォンに関する雑音のまとめ

また聴診器型 NAM マイクロフォンにて NAM のサンプリングを多数行った経験上，NAM サンプリングの際に NAM 信号以外に混入する，体表から直接センシングするという特殊性ゆえの様々な雑音をすでに既述のものもあるがもう一度整理しておく．

A) 機械雑音（ハムノイズなど）
B) 接触雑音（NAM マイクロフォンへの接触により起こる）
C) マイク側背部漏入気導外部雑音（マイクの構造上の欠陥）
D) 運動性体表雑音（歩行，随意運動，衣擦れなど）
E) 体内生理雑音（脈音，呼吸音など）
F) 体内伝導外部雑音（体を伝導して感音面から NAM と一緒に入る）

A)は DAT やコンピューターなどの収録機器の AC 電源が原因であったり，皮膚に接した NAM マイクロフォンとマイクアンプの電位差のために起こる．

B)は突発的に何かがマイクやそのコードに触れて起こる雑音であるが，もうひとつ別の原因で起こる重要な雑音でもある．NAM マイクロフォンを指で支持して皮膚に圧着することにより，一見静止しているかに見える指の震えにより起こる．

C)はマイクを改良する事で排除できる雑音であり，聴診器型からソフトシリコーン型になって大きく改善された点であり，気導音の排除が重要である．

D)については体伝導音の性質上やむを得ないが，自分の意志では制御不能の外部雑音と異なり，入力時には随意的にコントロール可能な雑音であるためあまり問題とならない．むしろ逆に非入力時の行動モニタリングや意図的雑音によるパラ言語表現入力として使用できる可能性もある．

E)について脈音はパワーが低く音声よりはるかに低域を占め，全く認識や通信の妨げとならない, 非発話時の呼吸音と NAM との鑑別は重要であるが, NAM 発話時には NAM そのものが呼気音であるため, NAM と通常呼吸音が重なることはなく，これも解決可能と考える．

F)が A)～E)までがすべて解決されたとしても残る体内を伝導する外部雑音である．体はローパスフィルターの役割を果たすとはいえ，低域の雑音は肉伝導する．感音面から NAM と一緒に入ってくるため，これだけは信号処理を駆使して除くべき雑音である．

つまり NAM マイクロフォンを製作するに当たって，雑音に関して製作者が責任を負うのは上の A)，B)，C)の三つである．

3.9.2　マイクアンプの工夫とハムノイズ対策

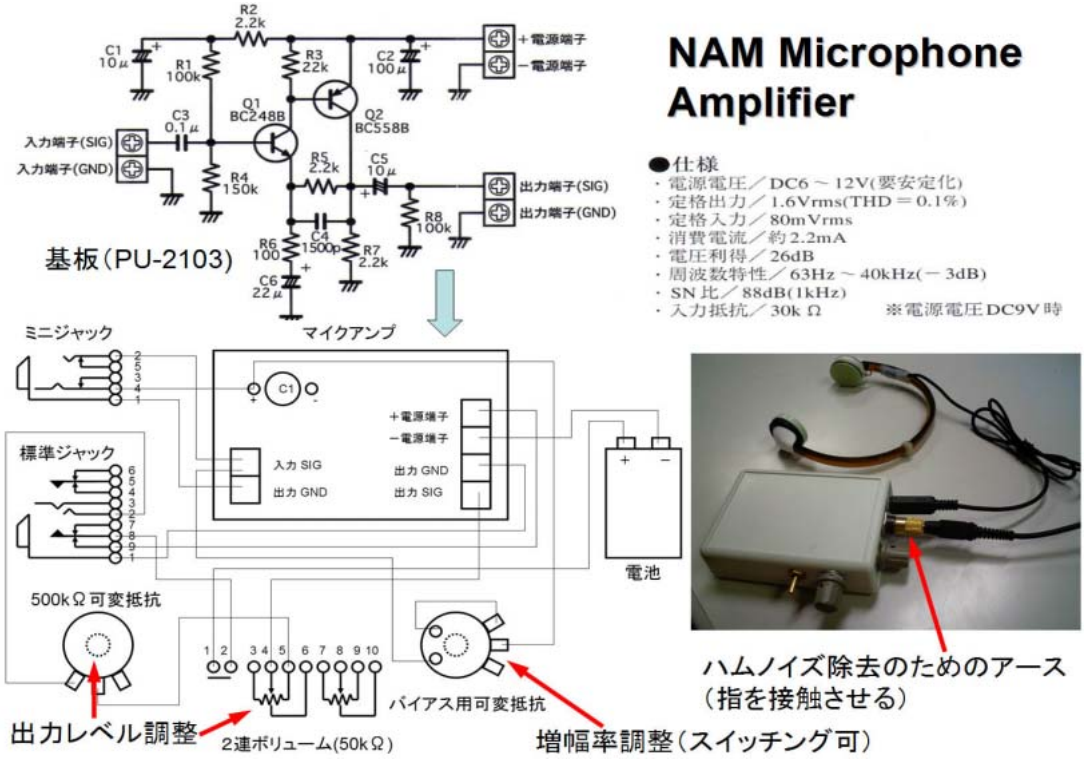


図 3.33　NAM マイクロフォンのマイクアンプ

体から離して使用する気導音マイクロフォンで音声を収録するときもハムノイズが経験されることがあるが，人間の体に直接接触させて使用するNAM マイクロフォンでは特にこの問題が顕著になる．現在の段階ではセンサー部分が試作段階であるため，センサー部からマイクアンプを 1m 前後のコードでつないで分離しているので，センサー部とマイクアンプ間の電位差がその原因であると考える．またセンサー部を小型にするため 3 極式でなく 2 極式コンデンサマイクロフォンを用いていることもその原因の一つである．

まず PC や DAT の AC 電源から直接入るハムノイズは，AC 電源を抜いてバッテリ使用することで大きく低減するが， PC などを机から離して膝の上に乗せたりすることも有効な手段である．しかしそれでも前述の電位差によるパワーの低いハムノイズは残る. 図 3.33 に現行マイクアンプの構造を参考のために掲げるが，マイクアンプ出力の標準ジャックを金属製にし，使用時にはここに指を触れて体とマイクアンプをアースしてやると，完全にハムノイズが除去できることがわかった. 図 3.34 がそのハムノイズ除去の例である．

将来的に NAM マイクロフォン内にマイクアンプやバッテリ，プロセッサー等が埋め込まれれば解決可能である．

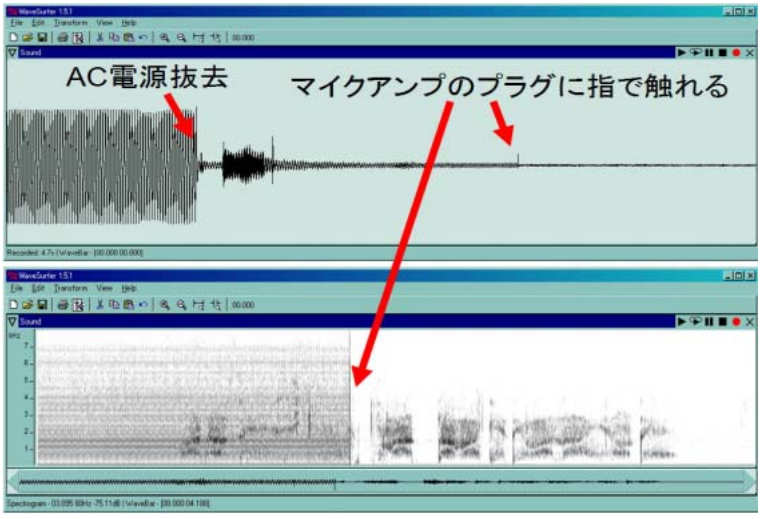


図 3.34　ハムノイズ除去の例

3.9.4　NAM マイクロフォンの固定法

聴診器型 NAM マイクロフォンでは，糊付き振動板が同時に固定の役目も果たしていた．実験時に糊付きプラスチックプレートを多く消費するのと着脱が煩雑なため，とりあえず実験時に手軽なように，ソフトシリコーン型は接着式にせず指で押しつけるだけの固定方式にした．

皮膚表面は肉眼的には平滑なように見えても，実は拡大してみれば非常に凹凸が多い．これは医療用超音波診断装置のプローブを皮膚に当てる際にも経験されることであるが，いくらペーストを塗ってもある程度圧をかけて皮膚表面を平滑化し，凹凸が原因ではさまれた微量の空気を排除しないとまったく画像が観察できない．NAM マイクロフォンも指で押さえるその圧を変化させると，感度は変化し，サンプリングされる信号も微妙に変化する．しかしフォルマント描出に最も適正な軽い圧をかけると，信号全体に低いサイクルの定常なノイズが乗ることがわかった．3.6 で視覚的簡易評価の際にはこの指で固定した NAM マイクロフォンを用いているので, 感度の高い NAM マイクロフォンほどこのノイズが現れ，S/N 比が低くなっている．この雑音そのものは 50Hz 以下であり，低周波すぎて認識や聞き取りにはさほど影響を与えない．

人間の筋肉はその筋肉の関与した部分が静止しているときでも，緊張していることがある．固定のためにも筋肉は使用され，たとえば指で何かを固定するときにも，指の伸筋と屈筋は共に緊張して固定のためのバランスを保っている．このとき肉眼では見えない振動が生じている．この振動により，指の表面とマイク裏面，またマイクの感音面と皮膚表面に微細な摩擦が生じ，これがこのノイズの原因である．

そこで指はまったく触れずに，女性のヘアバンド（カチューシャ）を切断して長さを調節し，図 3.35 のごとく NAM マイクロフォンを固定してみた．ヘアバンドはもともと適度な弾性があるので，熱でヘアバンドの曲率を調節すれば，NAM マイクロフォンに適度な圧もかけられる．右側の皮膚接触部

はあくまで固定用でありダミーである．この場合ネックバンドは体に触れておらずあくまで NAM マイクロフォンが耳下に固定されているのは左側の NAM マイクロフォンと右側の固定用ダミーの接触面の摩擦によってである．

図 3.35 の下段にネックバンド使用前（左），使用後（右）の NAM 信号を比較して掲示する．S/N 比の向上が視認できる．

図 3.35　ネックバンド式 NAM マイクロフォン

ただしこのネックバンド方式は，あくまで実験室用であって，静止してデータをサンプリングするためのものであり，当然首の上下運動や回転運動で運動性体表雑音が入ることは言うまでもない．しかし自分の頭部にマッチしたネックバンド式 NAM マイクロフォンさえ作れば，ハムノイズがなく、指の接触雑音が入らないきれいな感度一定の NAM 信号を、誰もが安定して実験室レベルでサンプリングすることが可能となった。

図 3.36　耳掛け式（補聴器方式）NAM マイクロフォン

聴診器型と同様，皮膚に粘着する方式は，接着するための物質選びが難航した．脳波電極用接着剤，熱冷まし用ゲル，入れ歯固定ゲル，付け睫用接着剤，液体絆創膏，頭髪用固定ゲル，脱毛ゲル等使用してみたが，シリコーンのごとき高分子素材と人体の皮膚のどちらにも強力に粘着してマイク重量を支えることは難しかった．しかもソフトシリコーンと皮膚との音響インピーダンス的連続性を絶ってしまっては元も子もない．現在は接着に軟らかく粘着性をもった封止用シリコーンゲル（旭化成 WACKER）を使用している．いくぶん粘着性は弱くこのままでは落として紛失の危険があるが，はずしてもそのままま貼り付けて使用できる利点があるのと，接着方式にして圧をかけなくても比較的良好な NAM 信号が得られる．

図 3.36 は海外製品の携帯電話用耳掛け式無線ヘッドセットのマイク部を改良して作成した，補聴器式の NAM マイクロフォンである．この NAM マ

イクロフォン表面に，上記の粘着性シリコーンゲルを用いると，もしマイク部がはずれても紛失する心配がなく安全である．粘着性ソフトシリコーンはあらかじめシートを作っておき，切って両面テープのように使用する．

また同じ耳掛け式でも，接着剤を用いず，図 3.37 のように，耳介の弾性とゴム被膜の摩擦とによって耳に固定する方式のものも作成した．NAM マイクロフォンに必要な軽度の圧もかけることが可能である．個人の耳介近傍の解剖学的形態への微調整が必要だが，一旦作ってしまえば，装着は最も容易である．現在この形態のヘッドセットで Bluetooth 無線とバッテリを組み込んだ軽量のものが実際に海外で販売されているので，NAM マイクロフォンでも製品化は容易と考える．

図 3.37　耳掛け摩擦圧着方式 NAM マイクロフォン

この耳掛け式は形態上，やがて登場するウェアラブルコンピューターなどの眼鏡型モニタ付属の NAM マイクロフォンにそのまま通ずる．認識において最適サンプリング部位が頭頸部の中間地点の，眼鏡の柄の終点からの延長上にあることから，実生活に日常的に用いられている眼鏡やサングラスに組み込んでしまったり，または着脱方式にするのは極めて実際的である．

図 3.38　眼鏡式 NAM マイクロフォン

市販のスポーツサングラスを用いて眼鏡式固定法を試作した（図 3.38）. この方式はネックバンド方式などと異なり首を動かす事による雑音が入らない極めて現実的な固定法である. 図 3.39 は市販の音声認識ソフトウェアに付属の USB ヘッドフォンに，NAM マイクロフォンを取り付けたものである.

図 3.39　ヘッドフォン式 NAM マイクロフォン

3.9.5　現行 NAM マイクロフォンの構造

実験，サンプル収集，デモなどに使用している現行 NAM の構造を図 3.40 に提示する．集音効果と外部雑音遮蔽の両方の意味で，パラボラ型の錫金属板を OCMSS 型にも OCWSS 型にも用いている．一次防音としてハードシリコーン素材を，外殻の二次防音素材に硬質のレジンを用いた．パラボラ型金属板に外殻のレジンから直接振動を伝わりにくくするためと，導線からの振動伝搬を防ぐため，パラボラ型金属板と導線は，可塑性があって振動を吸収しやすいエポキシ樹脂（接着剤）で支える形としている．媒体のソフトシリコーン表面には粘着性の透明ソフトシリコーンシートを貼付する．これらの NAM マイクロフォンに，実験室用途として用いるときはネックバンド固定方式を採用している．

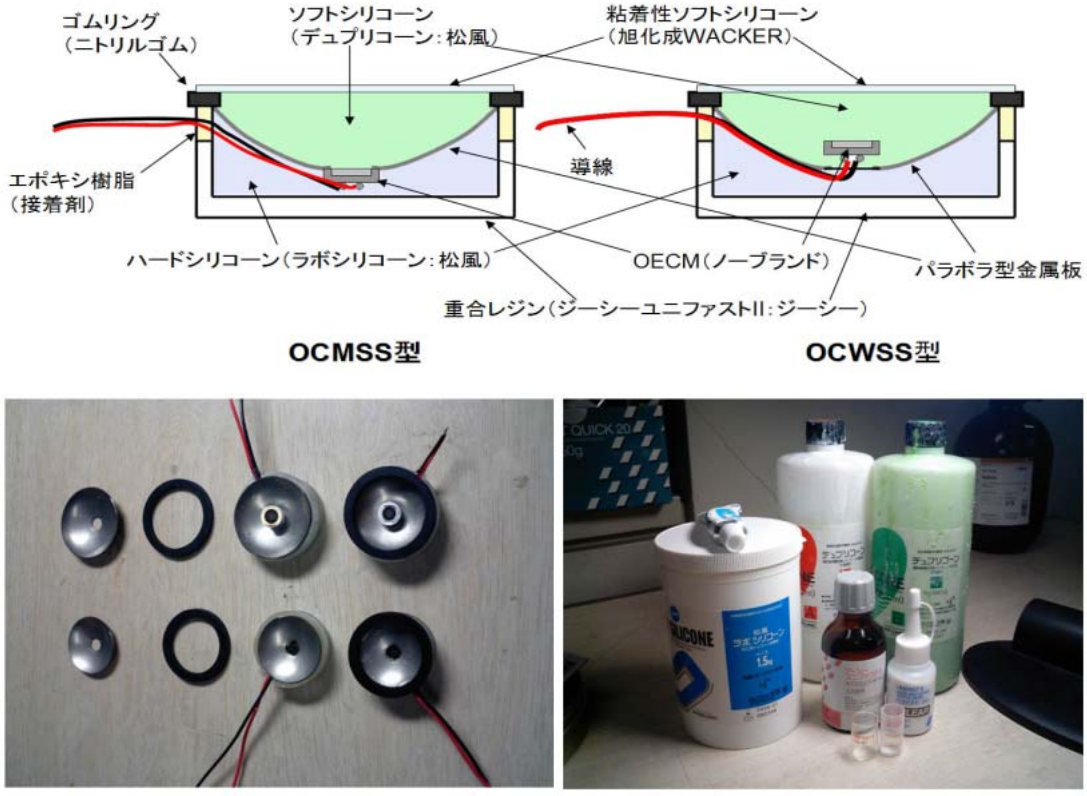


図 3.40　現行 NAM マイクロフォンの構造とその素材

3.10　他の接触型体伝導音センサーについて

体伝導音声を体表に接して収録するためのセンサーは NAM マイクロフォンだけではない．日本には「骨伝導マイク」と称される（マイクに「骨伝導」と言う言葉は意味がないということは序章で述べた）商品としての体伝導音センサーがあるし，台湾製の Throat マイクと呼ばれる商品がある[18]．日本でも最近 NAM マイクロフォンを原型にした肉伝導マイクロフォンが試作品として展示会などでデモされている．電子聴診器はその用途としての性質上音声よりもかなり低い周波数に感度を集中させているのでここからははずすとして，これらの体伝導音声センサーに共通していることは，雑音環境下での「通常音声収録が目的」ということである．当然感度は通常音声の収録に特化してあり，そのままでは NAM は収録できない．通常の使い方よりも信号を増幅して使えば NAM を収録できるものもある．ただその音質はまちまちであり，聞いてみた印象からは通常音声ほどには一般に NAM の収録に適しているとは言えないようである．

あくまで参考のため図 3.41〜図 3.44 に各センサーの特性の視覚的簡易評価を，前述の 3.6 での方法と同様に帯域，感度，雑音耐性に分けて簡易評価し，センサーごとにまとめて掲示する，NMHF の周波数応答は文字が小さすぎて見えないが，発話時の信号と TSP 信号の振幅を比較すれば，大まかなところはわかる．

骨伝導音マイクでは，音質を良くするために気導音も同時に収録しているものがあるが，骨伝導音だけを拾うものを選んだ．Throat マイクは添付の解説書を読むと，マイク内にすでに増幅回路が埋め込んである．取り外せないので，そのまま使用した．マイクアンプは条件を一定にするため，すべて今まで使用してきた自作のマイクアンプを用いた．対照に，標準的な現行ネックバンド式ソフトシリコーン型 NAM マイクロフォン（OCMSS 型）を用いた．

NAM マイクロフォンは，NAM の認識や聴取を前提に，NAM を効率的に高音質でサンプリングするためにもともと設計されたものであり，その基本姿勢は大きく異なっている．BTOS の収録はあくまで従である．

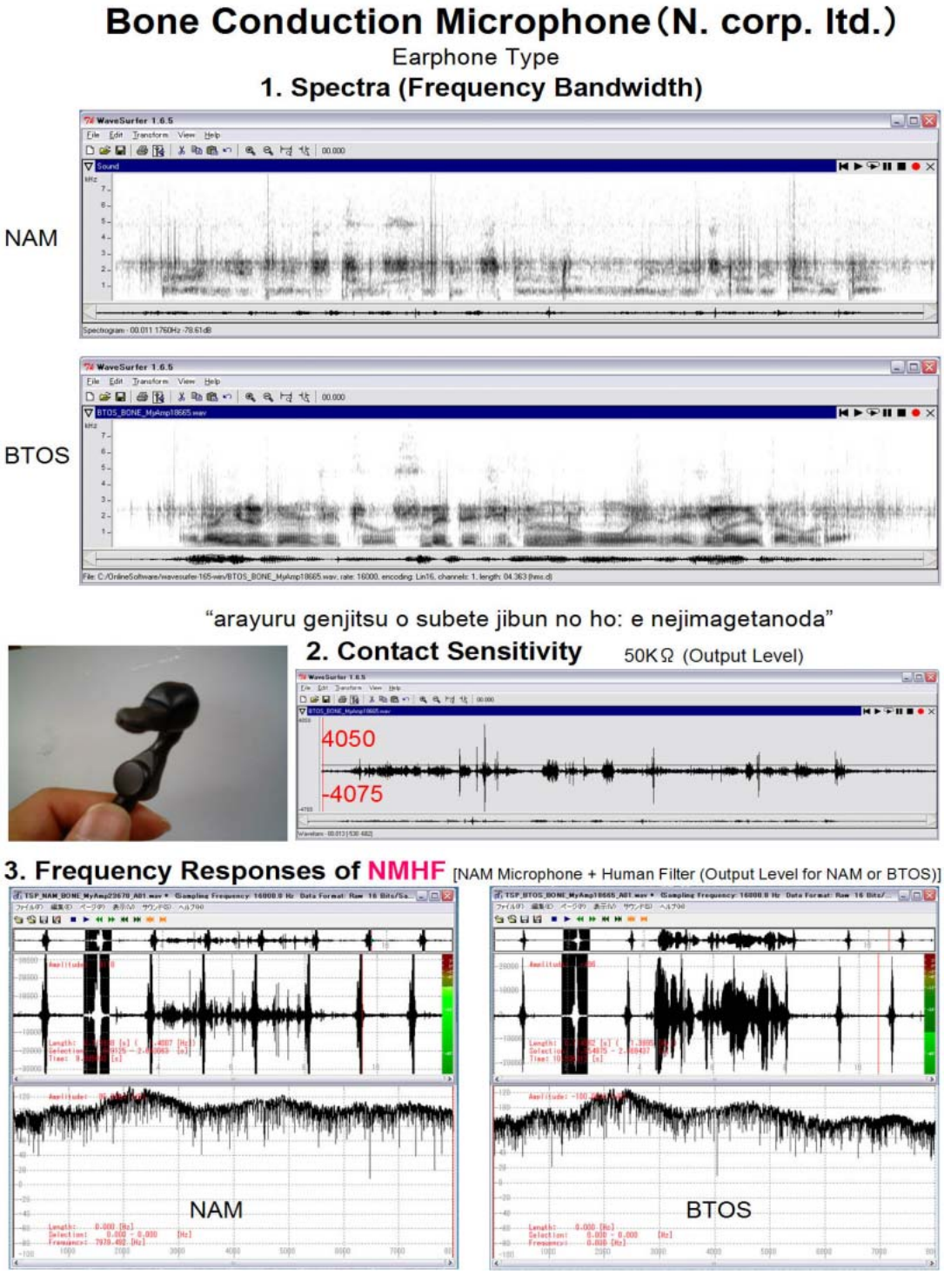


図 3.41　市販 N 社性骨伝導マイクロフォン

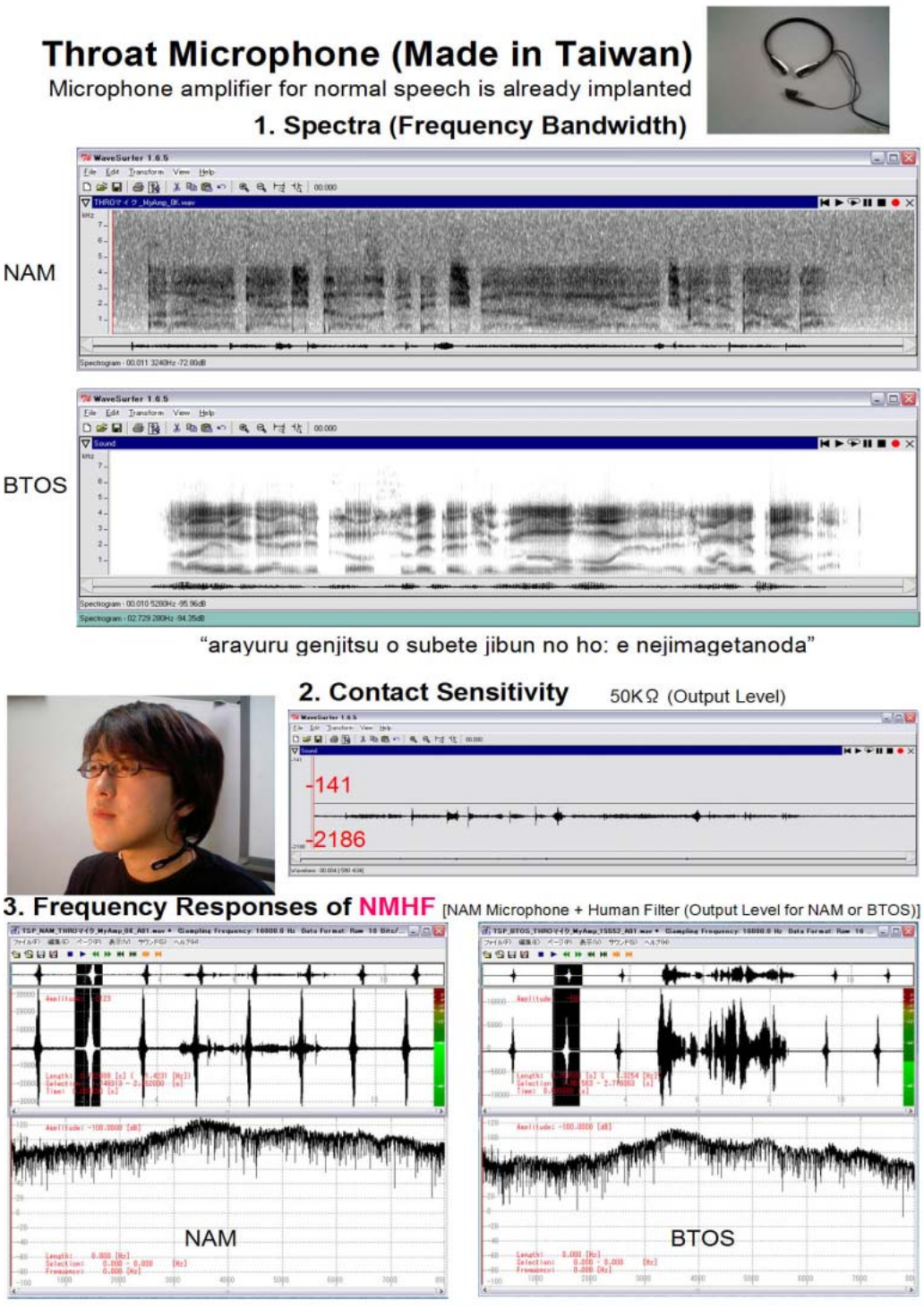


図 3.42　市販台湾製 Throat マイク

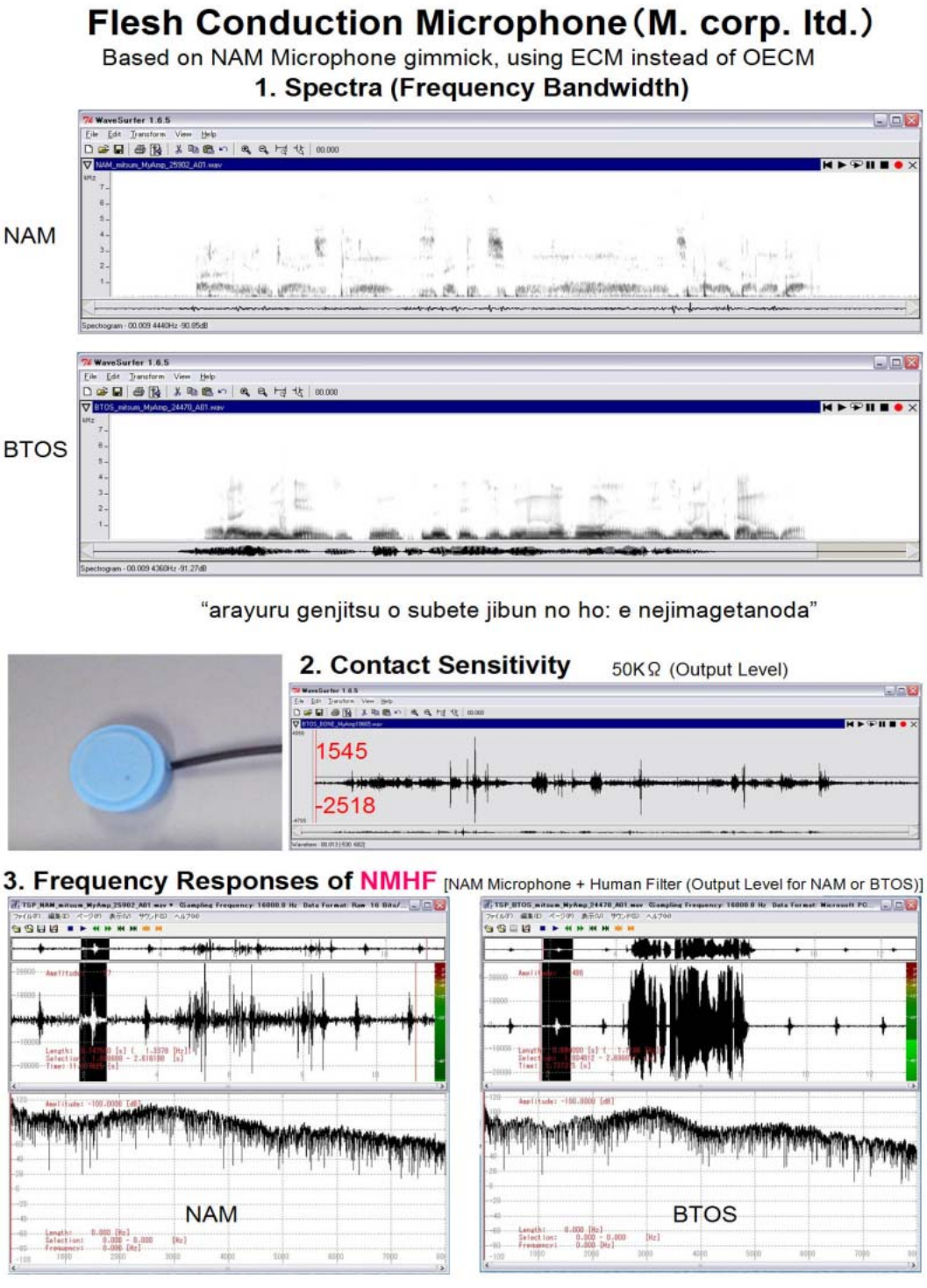


図 3.43　M 社製肉伝導マイクロフォン試作品

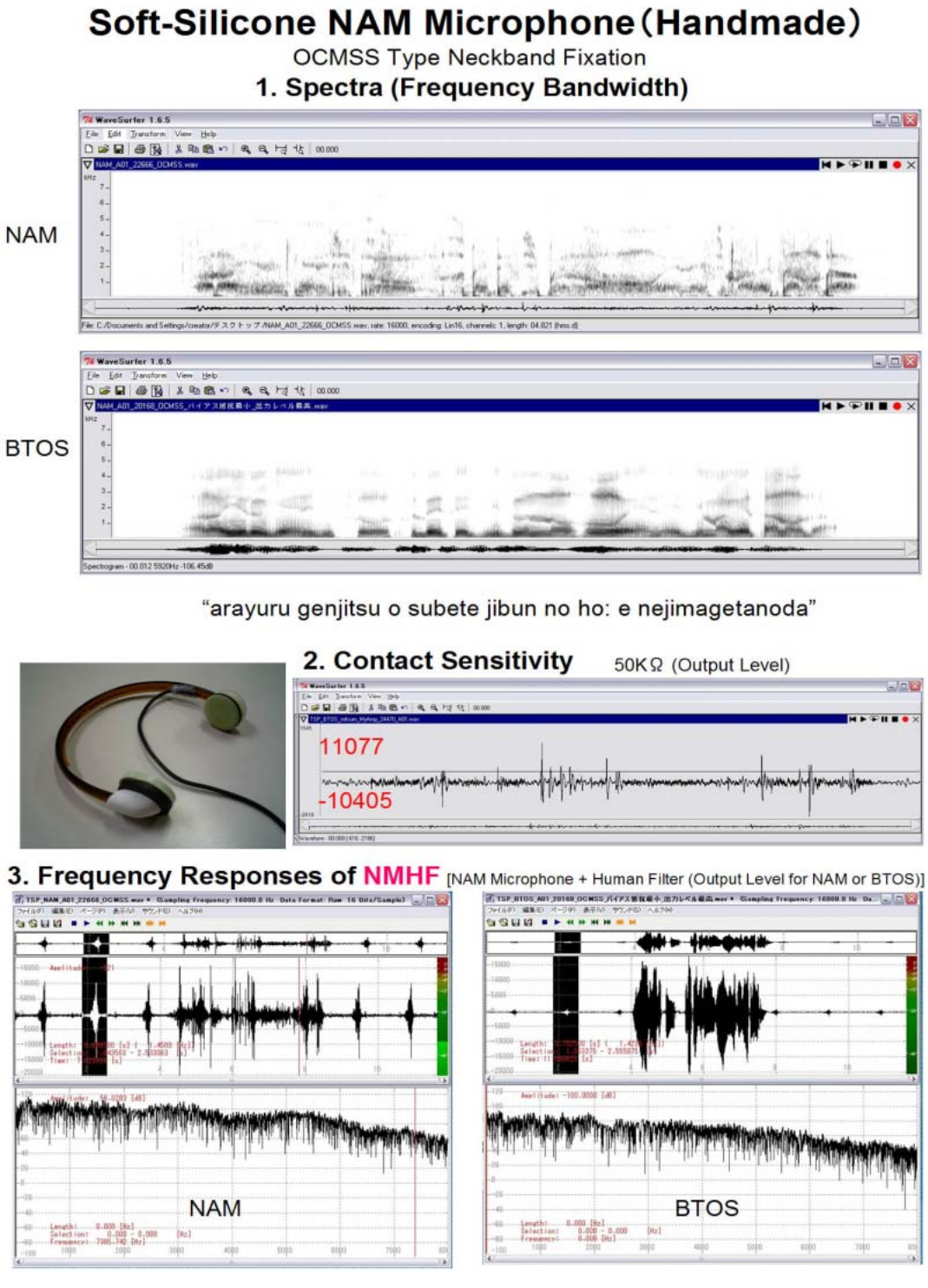


図 3.44　現行ソフトシリコーン型 NAM マイクロフォン（OCMSS）

3.11　同発話での気導音声・肉伝導音声の比較実験

NAM は「非可聴つぶやき」というが一体どれほど「聞こえない」のか？ どれくらいの大きさで発話すれば NAM と言うのか？　本当に口元のマイクでサンプリングするより感度が良いのか？　口元で何 dBA，NAM マイクロフォンで何 dBA なのか（これは計測が極めて難しい）？　NAM と言っても本当はささやき声くらいの大きさでしゃべっているのではないか？　等々の質問をよく受けるし，またそういう疑問は当然のことである．

ソフトシリコーン型 NAM マイクロフォンの，聴診器型 NAM マイクロフォンに対する優位性をこれまで確認してきたが，標準的な現行ソフトシリコーン型 NAM マイクロフォンを選び，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比較して，NAM と「気導 NAM 発話音」がどの程度異なるのかを客観的に確認する．

3.11.1　実験の方法

NAM マイクロフォ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フォンを製作しようとしても，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ーン型 NAM マイクロフォンとしては帯域，接触感度，雑音耐性について平均的な現行ネックバンド式 OCMSS タイプのものを選んだ．OECM を用いたものを選んだ理由は，気導音マイクロフォンとして，振動電極版を露出する前の同じコンデンサマイクロフォンを使用して比較するためである．マイクアンプは気導音と肉伝導音の増幅率と出力レベルを同等にして収録した．音のデジタル化は PC 内部のものを使用せず，外付けの USB インターフェースのサウンドカードを用いて，サンプリングレート 16KHz，16bit でステレオ収録した．気導音は NAM マイクロフォンに使ったのと同じコンデンサマイクロフォンを，口元から直接息がかからな

い程度に横にずらし 5cm 離した位置に固定して室内静音環境で収録した．

NAM はその発話者の置かれた環境によって音量が変わる．騒音環境では NAM も大きなパワーとなり，静音環境では，周りに聞きとられないためには，パワーの低いものとなる．したがって発話様式は，①NAM 概念の創始者である筆者が室内静音環境において最も NAM 発話らしいと考える NAM 発話と，②ささやき声発話，③通常音声発話である．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．

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，気導音雑音信号と音声信号のパワー比の違いを見るため，①と③について，実験者の前方 50cm の距離のスピーカーより TSP 信号の繰り返し音を雑音として出力しながら発話したものも収録した．

3.11.2　結果

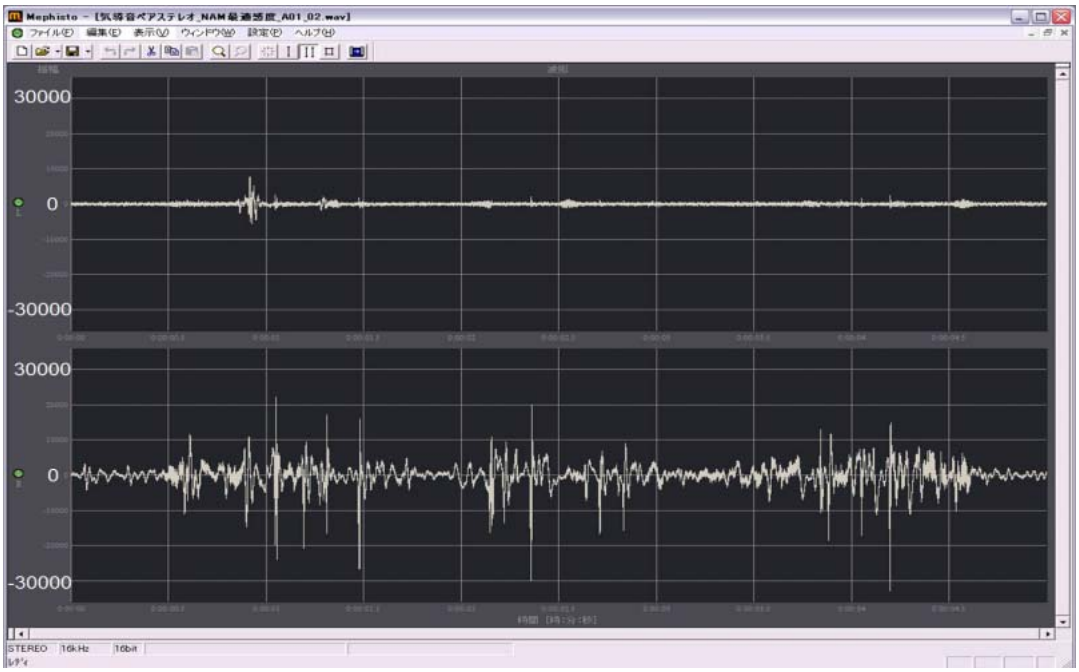


図 3.45　気導 NAM 発声音（上段）と NAM（下段）のステレオ収録

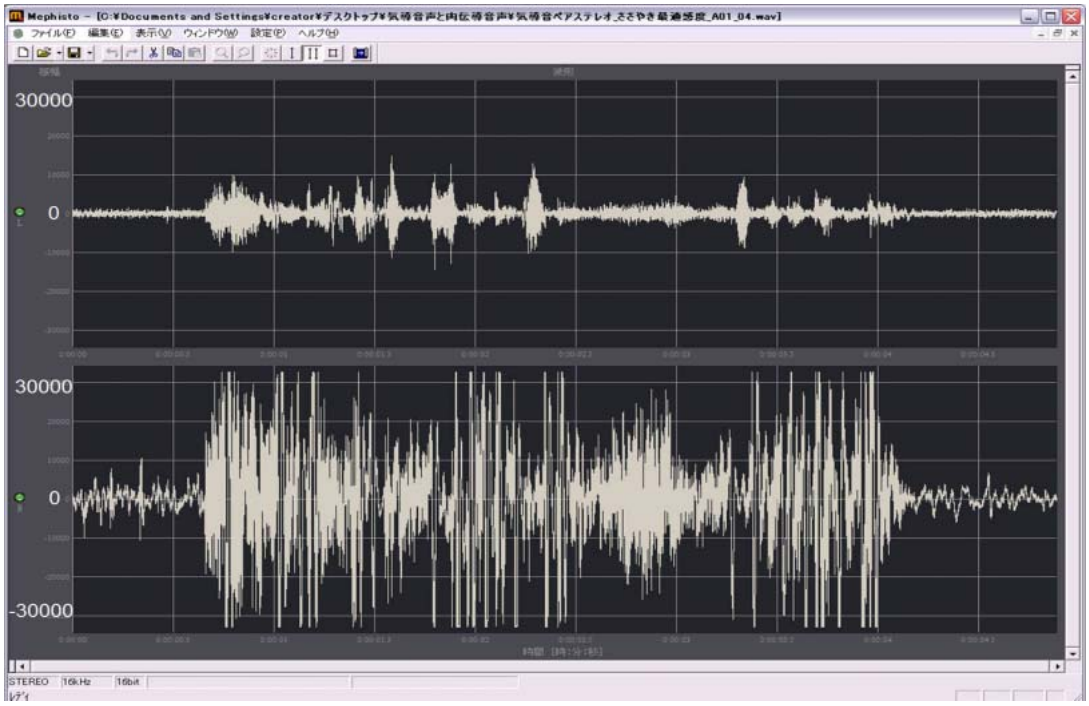


図 3.46　ささやき声（上段）と肉伝導ささやき声（下段）のステレオ収録

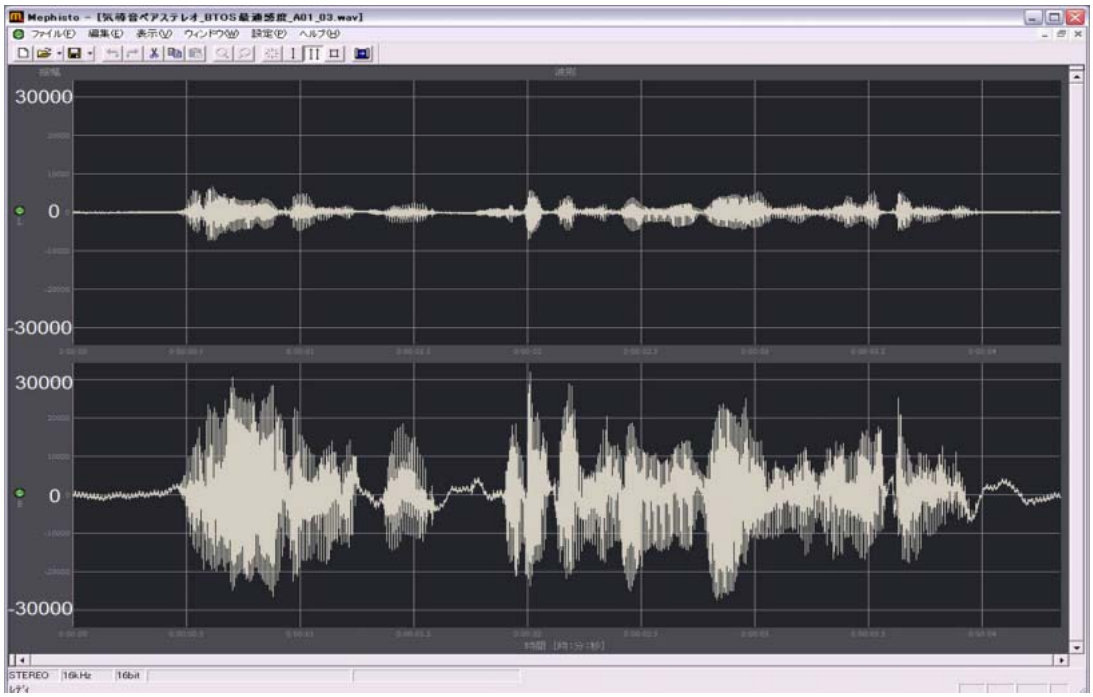


図 3.47　気導通常音声（上段）と BTOS（下段）のステレオ収録

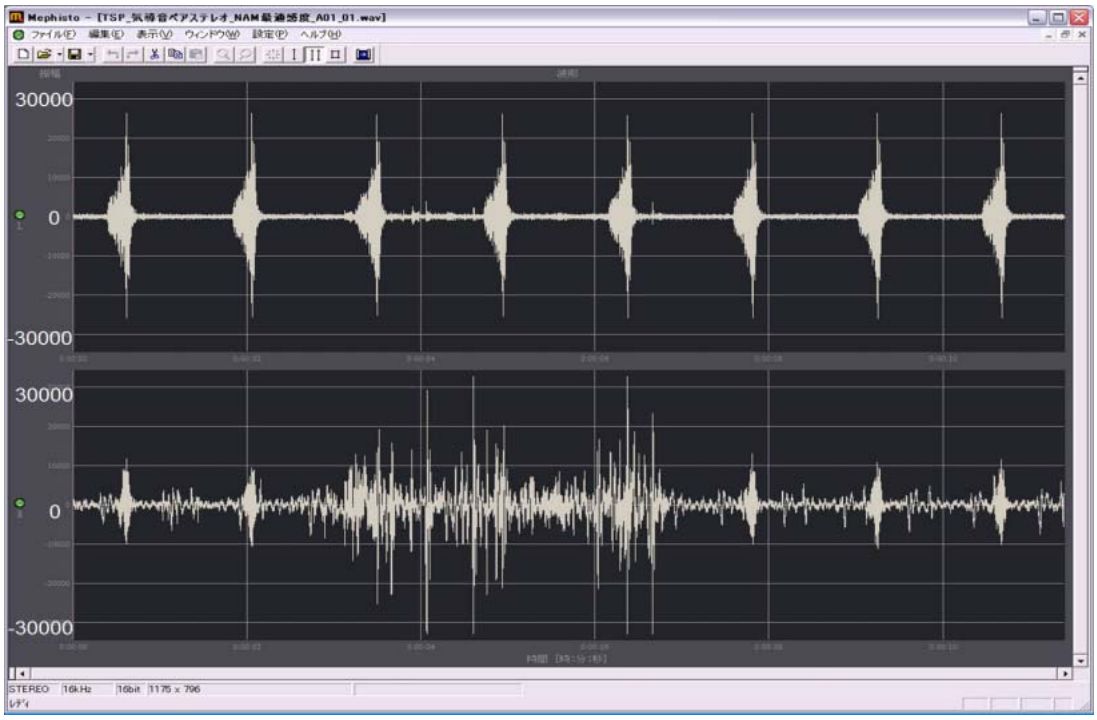


図 3.43　NAM 発話中の気導 TSP 信号（上段）と肉伝導 TSP 信号（下段）

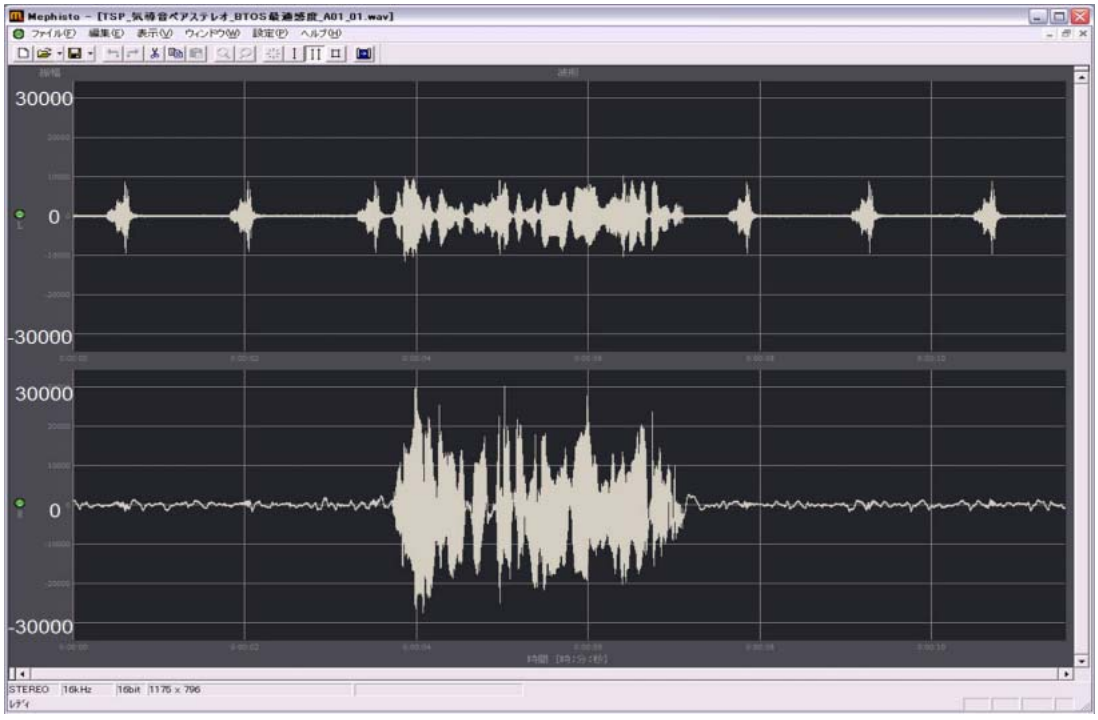


図 3.44　BTOS 発話中の気導 TSP 信号（上段）と肉伝導 TSP 信号（下段）

図 3.45 は気導 NAM 発話音つまり NAM 発話中に漏れ聞こえる呼気音と NAM マイクロフォ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，5.98dB であった．

図 3.46 は日常 1〜2 メートル先の相手に情報を伝えるために自然であると思われるささやき声の気導音と，その肉伝導音である．マイクアンプの出力レベルは気導音のささやき声に合わせた．

図 3.47 は気導音としての通常音声と，NAM マイクロフォンから収録される肉伝導の通常音声 BTOS である．マイクアンプ出力は 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，10.18dB であった．

図 3.48 は TSP 信号を背景雑音として繰り返し流しながら 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見にくいが，TSP 信号と音声信号との振幅を比較すれば，気導音環境と肉伝導音環境での目的信号と雑音信号の比率がわかる．

図 3.49 は TSP 信号を同音量で背景雑音として繰り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． BTOS の場合 TSP 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．つまり気導音の世界では，通常音声に近い音量を持つ TSP 信号が，NAM マイクロフォンによる肉伝導音の世界では BTOS に音量を合わせると「聞こえなく」なってしまう．

以上，同じコンデンサマイクロフォンを，一方はそのまま気導音マイクロフォンとして，もう一方は NAM マイクロフォンとして加工して使用し，同時にステレオ収録して，同じ発話様式の，気導音としての聞こえ方，肉伝導音としての聞こえ方を比較した．NAM でも BTOS でも目的信号である音声の感度は NAM マイクロフォンを使用した肉伝導音の場合の方が，気導音より高い．

参考のために図 3.45 と図 3.47 で NAM マイクロフォン側から収録された NAM 信号と BTOS 信号のトラックのスペクトラムを図 3.50 に掲げておく.

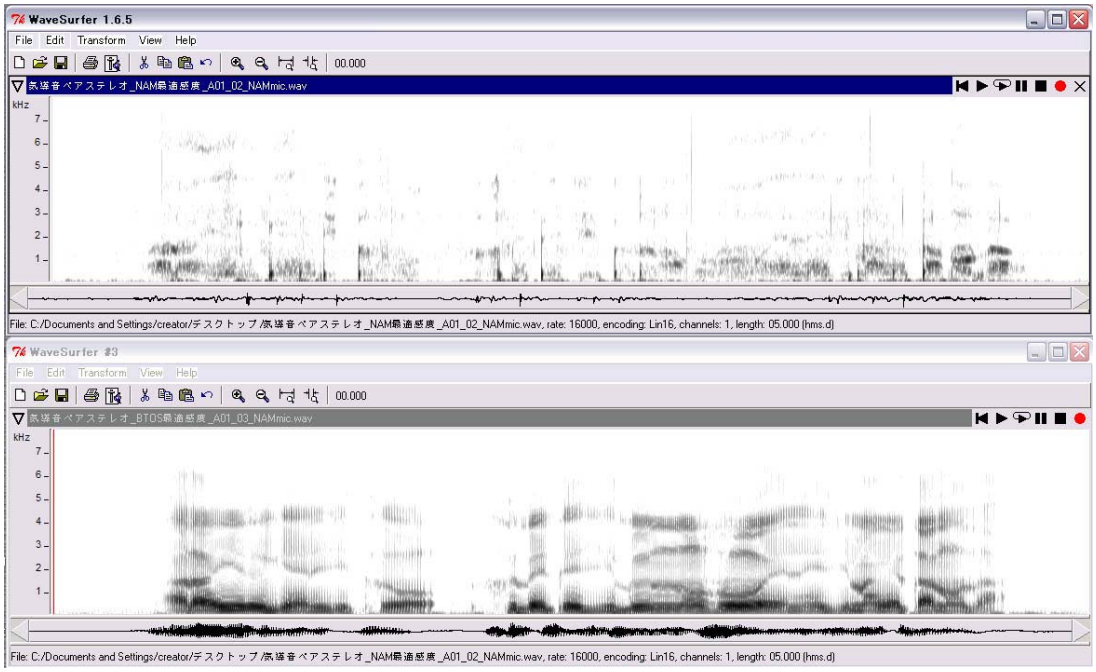


図 3.50　肉伝導音のトラックからの NAM と BTOS のスペクトラム

3.12　まとめと課題

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，帯域上限 2KHz だった聴診器型に比し，より広い帯域を持ち，HMM の話者適応による認識の単語認識精度を向上させた．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やすくなった．聞き取りテストでも，文章，意味単語，無意味単語のすべてにおいて，聴診器型に比し高い聞き取り率を示したが，無意味単語においては聞き取り率の低さが対照の気導音声に比して目立った．BTOS 音についても聴診器型に比べて聞き取り率の情報が見られ，しかも通常音声の聞き取り率にかなり迫っている．これにより NAM 音，BTOS 音自身や，簡単な信号処理に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．または NAM 音声に声質変換や音源付与などの処理を加えて通常音声化することで，今回聞き取り率の低かった無意味単語，未知語への聞き取り率を上げる課題をクリアすれば，周囲の状況に気兼ねしない，周囲雑音の影響も受けにくい，いわゆる声を関係ない周囲に振りまかない「無音声電話」などへの応用が期待できる．

新 NAM マイクロフォンは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．NAM マイクロフォンの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もの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内部設置方法を検討すること，また皮膚からの音媒体として，素材の種類と硬さや弾性を変えての実験も必要である．場合によっては内部のセンサーそのものの設計を肉伝導音用途に改造する必要があるかも知れない．

また同時にこういった接触性の体内伝導音マイクロフォンの評価のための，規格化された定量的特性測定法を論議し，コンセンサスを得ることと，また安価な測定機器の普及も望まれる. 今回の視覚的評価でその原型を提案した.

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，Bluetooth などの規格化されたデバイス間無線通信も関連企業にて検討中である.

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課題である．現在の気導音マイクロフォンのように気軽に自由に市販された NAM マイクロフォンを音声信号処理の専門家が手にすることができるようになれば，各自の専門性を生かして可能性は大きく広がり，このコミュニケーション法の確固たる基礎が想像するより早く出来上がると考えるからであり．また老若のユーザーによる思いも寄らない使用法が発見されると思うからである

第4章 縦アレイ NAM マイクロフォンによる韻律表現

4.1 はじめに

ソフトシリコーンを音媒体に用いた第二世代 NAM マイクロフォンを開発することで，NAM 認識（いわゆる無音声認識）の認識率が向上しただけではなく，NAM の帯域の拡張，その音質の飛躍的に向上により周囲の環境に気兼ねしない無音声電話の実現の方向性も見えてきた．それらを包括する NAM Interface Communication という概念を実現化するにあたって，避けては通れない問題がある．韻律，特に基本周波数（ピッチ）の問題である．

無音声認識においては，認識にピッチのモデルが必要であると考えられる中国語の問題がある．四声という音節内のピッチ変動で言葉の意味が全く変わってしまう中国語は現在 12 億人に話されている．また日本語でもピッチモデルを用いた認識が登場する可能性がある．

また無音声電話においては声質変換などで通常音声化する際に必要なパラメータとしてピッチ情報が不可欠である．通常音声化されたものが自然に聞こえるかどうかは，与えられるピッチ情報にかかっていると言ってもよい．現に個人性まで忠実に再現して見事に通常音声化された NAM を聞いたとき，その予測して与えたピッチ情報が単純なものであると，逆に声としての品質が良いだけに，対話音声としての不自然さが目立つようになるからである．

BTOS は声帯振動が音源となるので，ピッチは抽出可能である．しかし雑音駆動の無声音である NAM では，自己相関法などによるピッチ抽出は不可

能であり，F0 曲線はプロットされない.

収録した NAM 音サンプルを聞いてみると，韻律は存在し，ピッチらしきものも感じ取ることができる. 標準語で読んだ NAM と関西弁で読んだ NAM の違いも判別できる．しかしそれは単にアクセント情報のみから聞き手の脳が判断しているだけかもしれない． NAM 音サンプルのみからピッチを予測する情報を取り出す方法は，それはそれで大きなひとつのテーマとなりうるが，実際に今のところまだ基本的なアイデアも手法も確立されていない.

この章ではこの NAM マイクロフォンを同側で縦に２つアレイ化して装着し同期ステレオ収録することにより, F0 とは違った視点で発話のピッチもしくは韻律意図を検出・表現する方を提案し，NAM のピッチを予測する可能性のある方法を紹介する．またこの方法で肉伝導音声をステレオ収録したときに広がる, 肉媒体の音場としての人体という魅力的なテーマを提示したい.

NAM マイクロフォンを用いてサンプリングした肉伝導音声は，口からの放射特性などの情報を欠いていると思われる．しかしここで考えてみれば NAM マイクロフォンはもともと体に密着して音をサンプリングするため，デバイス中にいくつものマイクロフォンを仕込んでもいいわけである．音声の発生源に近接して，複数のマイクロフォンから複数の NAM データが得られるという利点もあるわけである．そこには単独音声ファイルとは次元の異なる多くの情報も得られる利点があることはマイクロフォンアレイを専門とされている方々には当たり前の事実であろう.

4.2　ピッチと喉頭部の上下運動

のど仏に手を当てながら低い声，高い声を発声してみれば誰にでもわかるが，一般に喉頭は声の高低に応じて上下に運動することは，古くから知られている．医療用超音波イメージング装置のプローブを図 4.1 のように前頸部正中に縦に当てれば，喉頭部の韻律変化に伴う上下動をリアルタイムに，秒間 30 フレーム以上の時間解像度で観察することが可能である.

図 4.1　超音波イメージング装置で観察する喉頭部の上下動
（コンベックスプローブを頸部正中線に沿って縦に当てる）

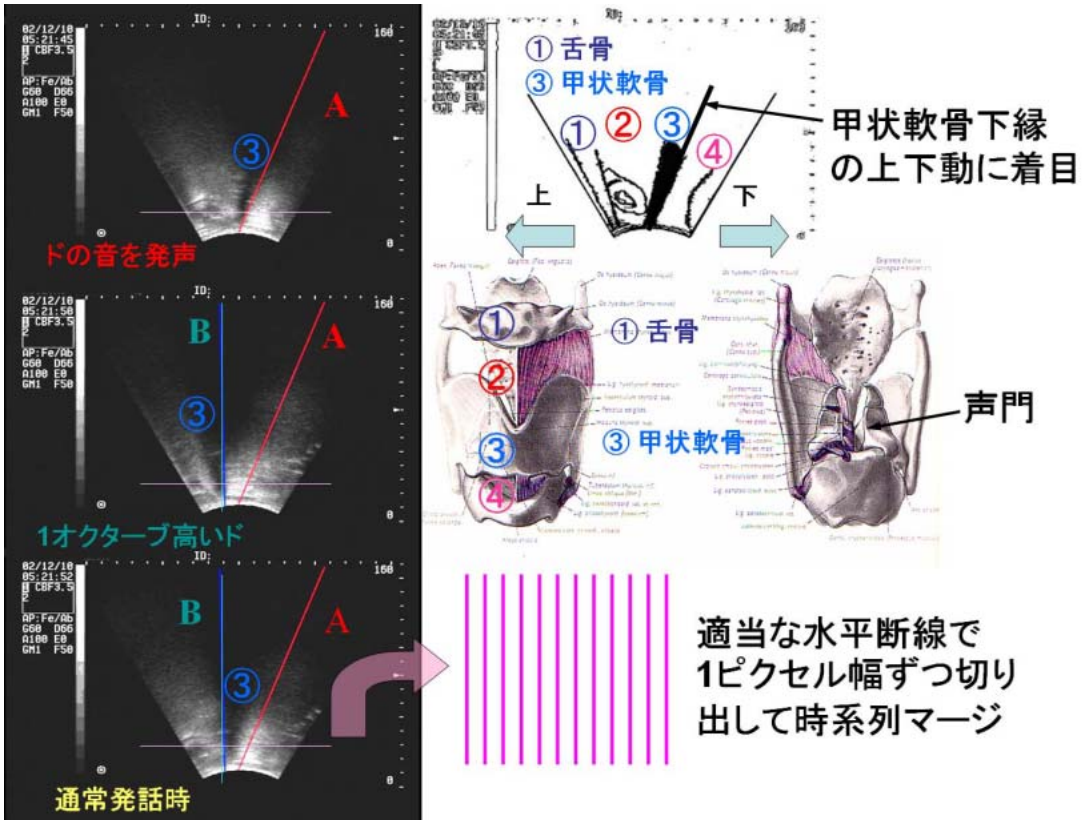


図 4.2　超音波イメージング装置による喉頭の上下運動の観察
（写真・模式図ともに左側が体の上側に相当．解剖図は人間の喉頭）

超音波イメージング装置では，図 4.2 の左の 3 枚の写真のように喉頭部の韻律変化に伴う上下動を，リアルタイムに観察することができる（写真の左が上）．白黒の動画として映るが，一般に骨や軟骨のある部分は，超音波の反射により黒い影となって表現され，筋肉，皮膚，結合組織などのいわゆる肉の部分は白く映って構造が描出される．図 4.2 の右中段の解剖図のごとく，喉頭には舌骨や甲状軟骨などの骨や軟骨があって，その全体の構造は観察が難しい. 特に声門部は甲状軟骨の影になってその構造を見ることはできない. しかしその上下動は，喉頭部全体が一緒に移動するため，基準さえ決めればどの程度の位置にあるかがわかる．ここでは甲状軟骨の下縁が最も白黒の境界が鮮明で観察しやすいため，図 4.2 の右上の模式図のごとく③と④の白黒の境界線である甲状軟骨下縁を上下動の基準線とする．

図 4.2 の左の上段の写真で個人特有の基本周波数に近い「ド」を発声したときの甲状軟骨下縁は線 A の位置に当たる．中段の写真で 1 オクターブ高い「ド」を発声したときの甲状軟骨下縁は線 B の位置まで移動する．通常の発話時にはこの甲状軟骨下縁が，下段の写真のように，この線 A と線 B の間を声の高低によって揺れ動く．この各フレーム画像を適当な水平断線で一ピクセル幅ずつ切り出し時系列マージすれば，喉頭の上下動がグラフのように視認できるはずである．

このようにして作ったのが，図 4.3 の下段の図である．上に音声波形や F0 曲線と同期させて表示してみる. まずキャリブレーションとしてド, ミ, ソ, ド（1 オクターブ高いド）を「a」の音韻で発声した．必ずしも絶対音階のドではなく，店内放送などのチャイムのメロディーを本人が発声しやすいように発声する．その後，通常音声，NAM ともに朗読発話内容は「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．線 A は「ド」の高さの声を発声するときの甲状軟骨下縁の高さで，ほぼその個人特有の自然発声時基本周波数の高さに相当すると考えられ，線 B はその 1 オクターブ上の「ド」の高さの声を発声するときの甲状軟骨下縁の高さに相当する．

甲状軟骨下縁の上下動を表す③の黒く表示された部分と④白く表示された

部分の白黒境界線の波形の揺れが，喉頭全体の上下動を表すと考えられ，これに LEI（Laryngeal Elevation Index）曲線という名前を付ける．有声音部において，LEI 曲線と図の上の音声データの F0 曲線との類似から，LEI 曲線が声の高さの一つの指標となることがわかる．

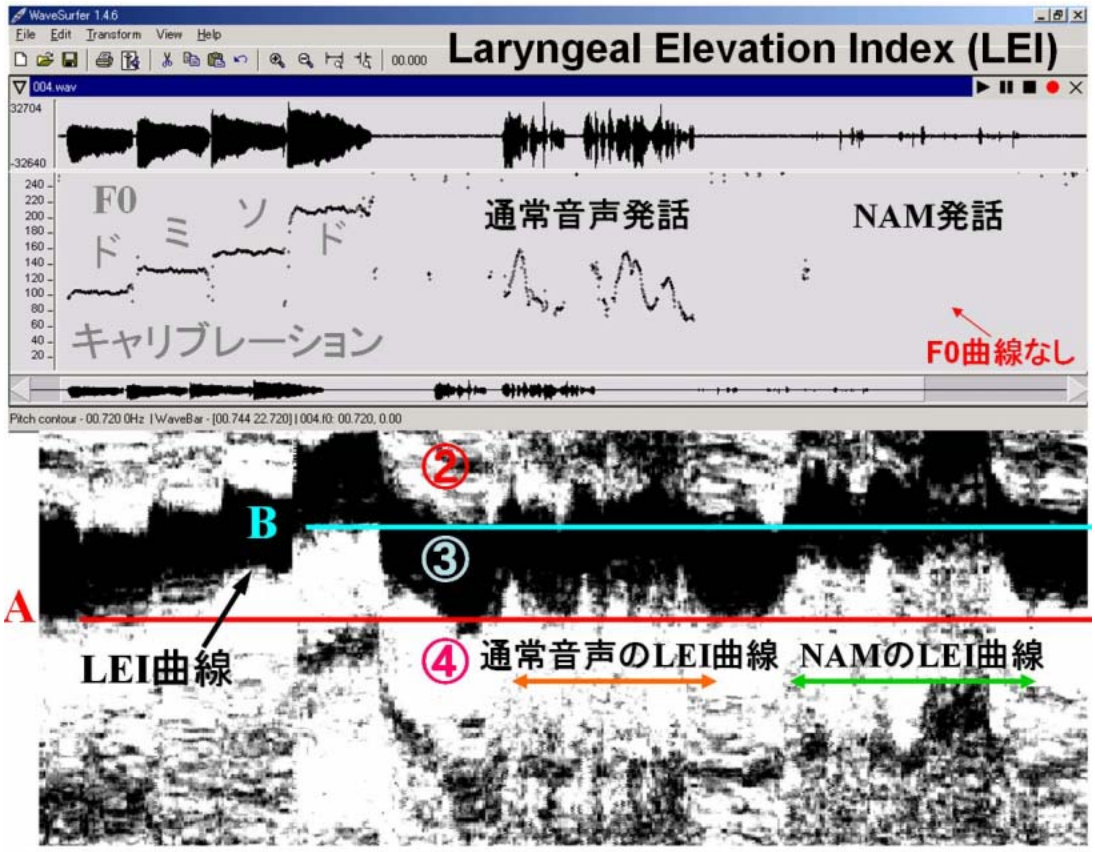


図 4.3　F0 曲線と LEI 曲線との比較

当然 NAM 発話時には上図の音声データからはピッチ抽出は不可能で F0 曲線は描出されないが，喉頭の上下動を表す LEI 曲線は描出されている．つまり無声音である NAM 発声時にも，あるピッチの声を出す準備状態として，通常音声と同様に喉頭の上下動を，通常音声よりやや高い声の発声準備状態として行っていることがわかる．

発話時の喉頭部の上下動が BTOS にも NAM にもあるならば, 喉頭の上下動を検知するデバイスを作成すれば, F0 とは別角度から見たピッチ予測ができることになる．NAM マイクロフォンとは別に頸部になんらかのセンシング装置を付ければ良いことになるが，デバイスが二つに分かれたものを，それぞれ体表に接着するのは実用的でない．できれば NAM マイクロフォンとしては単独デバイスで，サンプリングされた音情報からそれを検出することはできないものであろうか？　それを可能にするために考案した方法が NAM マイクロフォンの縦ステレオ化である．

4.3　縦アレイ NAM マイクロフォンの原理

喉頭部全体が声の高低に応じて上下動をするならば，音源となる声門も上下動をするはずである. それなら音源の移動範囲を上下二つのマイクで挟み, 縦にステレオで同期収録して移動音源定位をする方法を考案した．定位の方法は様々な手法があるが，一番簡単なパワー比を用いることにした.

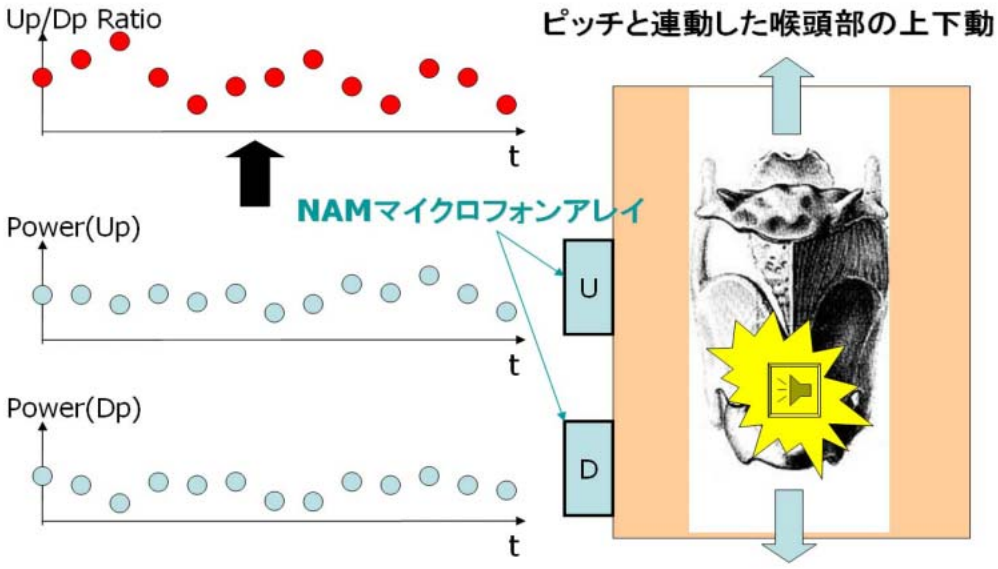


図 4.4　縦アレイ NAM マイクロフォンの原理

図 4.4 にその原理を模式化する．音源から大小様々なパワーの音が出ていても, 音源が上昇すれば相対的に上部 NAM マイクロフォン(U)のパワー(Up)が下部 NAM マイクロフォン(D)のパワー(Dp)に比して高くなり Up/Dp パワー比は上がる．音源が下降すれば Dp が Up に比して相対的に高くなるので Up/Dp パワー比は下がる．つまり喉頭の上下動に連動して Up/Dp パワー比が上下することになり，Up/Dp パワー比から喉頭の解剖学的な高さ，ひいてはピッチが推定できる可能性が高い．NAM のごとき無声音であっても，音源である乱流雑音は気道の最狭窄部である声門部で最強となり，その最強点が上下動をすると考えられる．

4.4　縦アレイ NAM マイクロフォンの方法

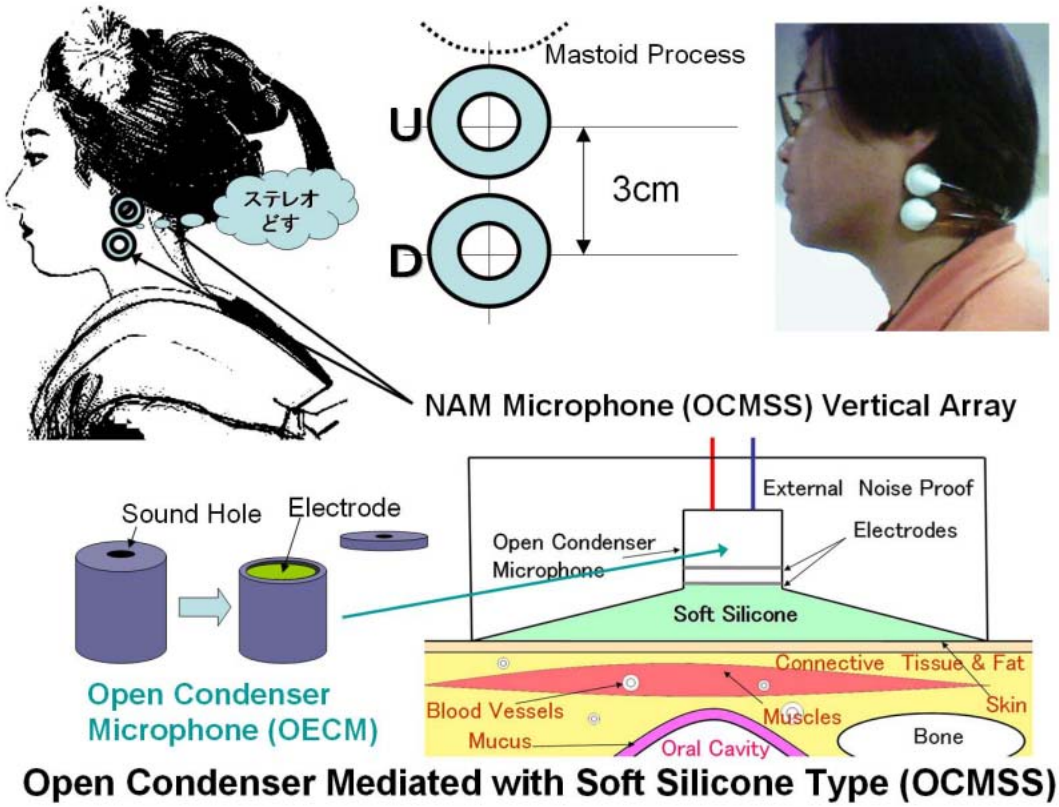


図 4.5　縦アレイ NAM マイクロフォンの方法

図 4.5 に選択した使用した NAM マイクロフォンの構造と，その装着方法を示す．NAM マイクロフォンとしてソフトシリコーンタイプの OCMSS タイプを選択し，これを音響学的に同じ構造に複製した．まず上部 NAM マイクロフォン(U)を，通常装着時と同じ乳様突起直下の耳介後下方部に装着し，胸鎖乳突筋に沿ってその 3cm 垂直下方に，下部 NAM マイクロフォン(D)を装着する．NAM マイクロフォンどうしは触れないよう考慮し，装着法はネックバンド方式を用いた. マイクアンプも同規格のものを複製し同じ増幅率, と同じ出力レベルに調整し，サンプリングレート 16KHz，16bit にて同期ステレオ録音を行った．ノート PC 内のサウンドカードは使用せず，ステレオ収録が可能な USB 接続のサウンド AD 変換デバイスを用いた.

上下マイクロフォンアレイで喉頭位置を定位するこの手法を SOL（Stereophonic Orientation of Larynx）法と名付ける.

4.5 結果

4.5.1 BTOS の Up/Dp パワー比

まず F0 曲線との比較が容易な BTOS 音よりステレオ収録を行った．もし BTOS 音で F0 と Up/Dp パワー比との間に相関が全くなければ, この上下ステレオ収録のパワー比をとる方法で，喉頭の上下動を推定し，その結果ピッチが推定できるという仮説そのものが誤りであったことになる.

最初のサンプルは BTOS 発声にて，すべて同音韻で，連続的に有声音ばかりで読み上げたものである.「ド，レ，ミ，ファ，ソ，ラ，シ，ド」の８音階を同音韻「a」で途切れなく発声した．図 4.6 がその上下 BTOS 音である.

図 4.7 は左上が BTOS 音の F0 曲線と（上部 NAM マイクロフォンのもの），左下がステレオ収録により得られた Up/Dp パワー比曲線である．右の図は横軸に F0 基本周波数（ピッチ），縦軸に Up/Dp パワー比をとったときの散布図である．相関係数は 0.869 であり，強い相関を認めた.

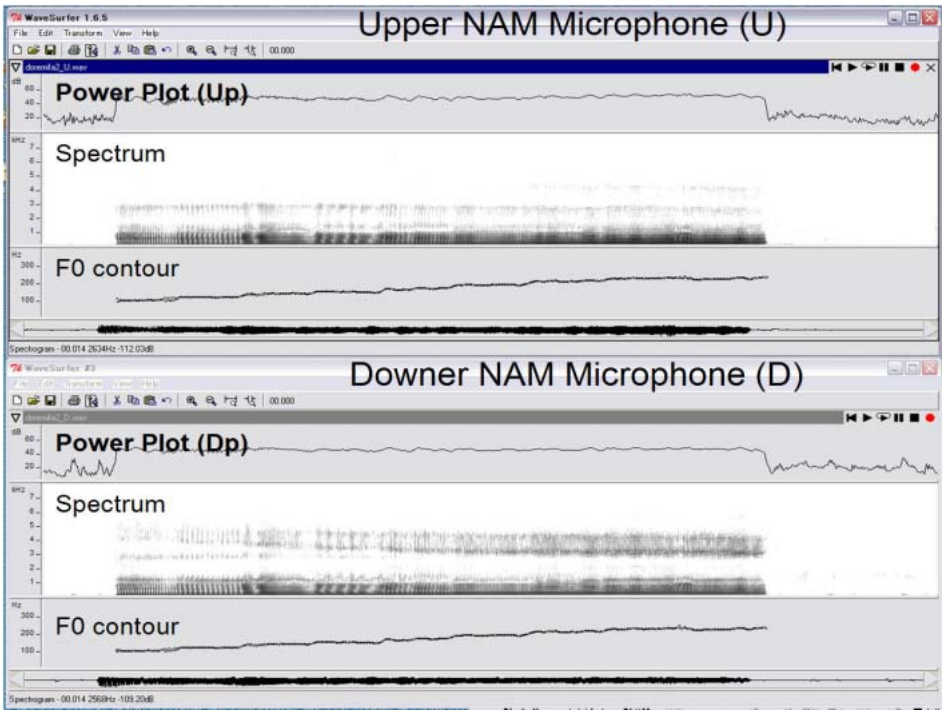


図 4.6　BTOS にて同音韻「a」の 8 音階発声

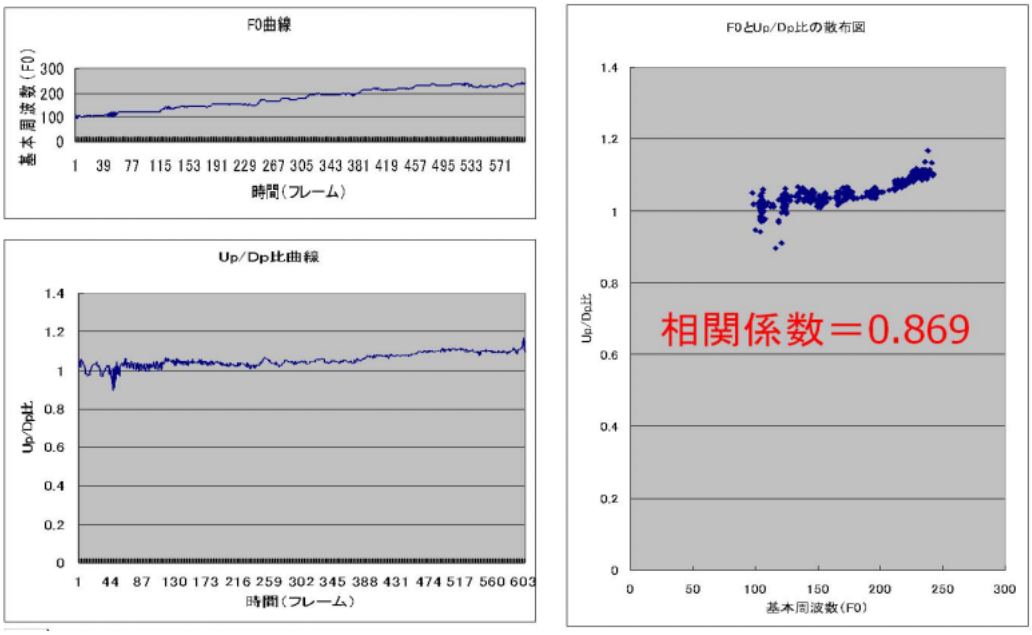


図 4.7　F0 と Up/Dp パワー比の相関

次にBTOSによる読み上げ文収録を行った.収録内容は最初にCalibrationとして「a」の音韻で「ド，ミ，ソ，1オクターブ上のド」を発声し，その後音素バランス文「一週間ばかり，ニューヨークを取材した」を読み上げた.

図4.8は収録した上NAMマイクロフォン（U）と下NAMマイクロフォン（D）のBTOS音信号，スペクトラムとパワー曲線である.

NAMの通常音声化の際に，F0に代わるパラメータとしてアクセント情報のパワー曲線を用いてはどうかという考えもあるが，図4.8のキャリブレーションのパワー曲線を見てわかるように，パワーのみでは1オクターブも高さの異なる音声が，ほぼ同じ値で表されてしまうことになる.

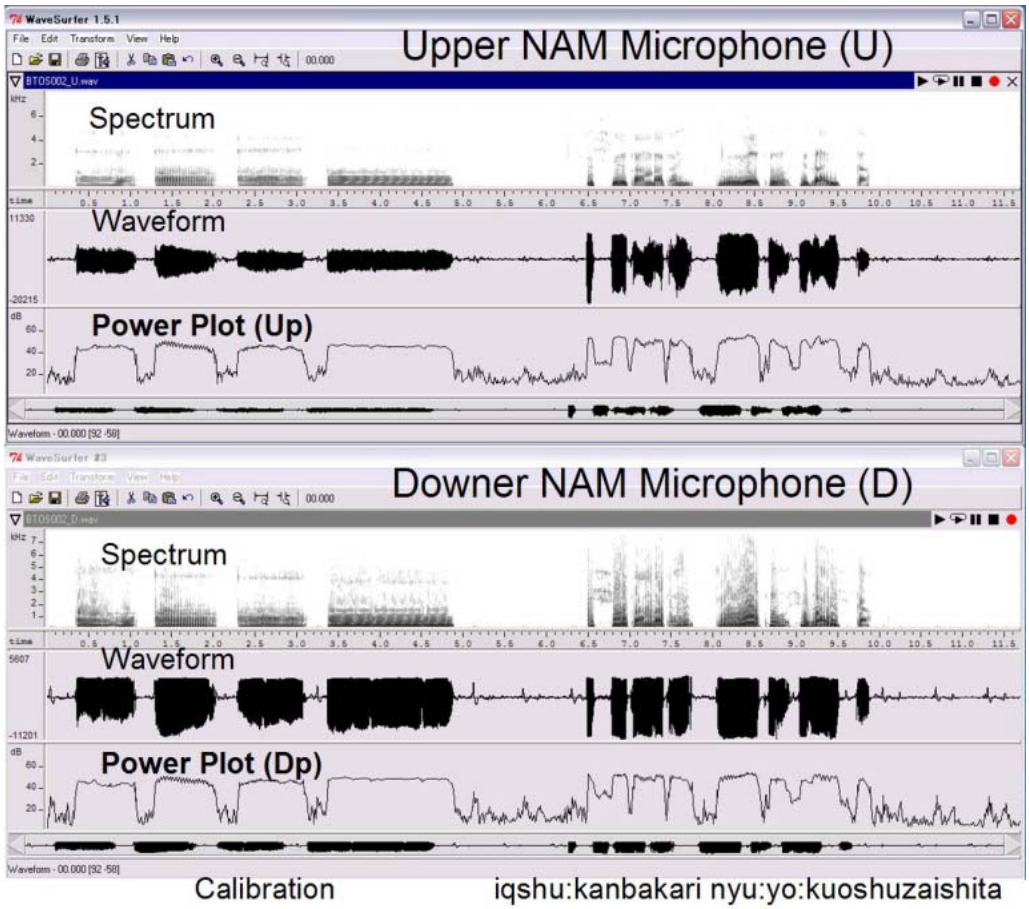


図4.8　縦アレイNAMマイクロフォンでステレオ収録した上下BTOS音

上部 NAM マイクロフォンでサンプリングされた今の単独信号を, 再度 F0 曲線も加えて表示する．BTOS は肉伝導音であるが，通常音声の気導音収録と同様に基本周波数が抽出され，図 4.9 のごとく F0 曲線が描出される．キャリブレーションで発声した「ド」と「1 オクターブ高いド」はその基本周波数においてほぼ倍になっていて，正しく音階は発声されていることが観察される．F0 曲線は有声音部のみに現れ，無声音部や無音部は描出されない．

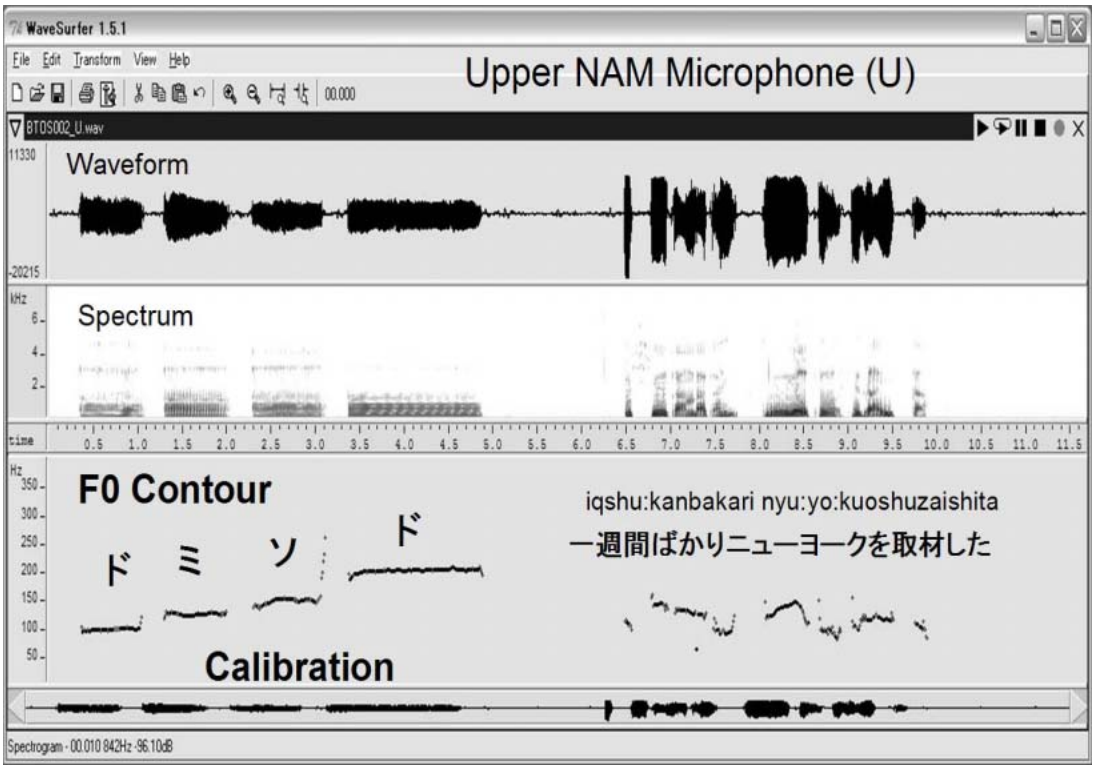


図 4.9　上部 NAM マイクロフォン収録 BTOS の F0 曲線

図 4.10 に上部 NAM マイクロフォンでサンプリングされた信号のパワーを Up，下部 NAM マイクロフォンでサンプリングされた信号のパワーを Dp としたときの Up/Dp パワー比を折れ線グラフ化したものを掲げる．同時に単独 NAM マイクロフォンの音信号より得られた前図の F0 曲線の部分を，比較のために時間軸を重ねて下に並べてみた．

まずわかることは，無音部では小さなノイズのため Up/Dp パワー比は極端値をとって大きく上下に変動するということ．また F0 のとぎれの部分に応じた位置が，比率 1 からはみ出して音素によって特定の特異値をとっていることもわかる．これは舌先や舌背など声門以外のところで音源の最強点を持つ無声子音に相当すると考えられる．例えば「sh」などの摩擦性の無声子音は口唇近くで発声するため U のパワーが相対的に高くなり，図 4.10 の矢印のように特異値をとる．

Up/Dp パワー比曲線から，これらの極端値，特異値を除いたものが，有声音部の喉頭の上下動を表す曲線に当たると考えられる．キャリブレーション部では相対的に F0 の高さに応じて，Up/Dp パワー比曲線が階段状に上昇しているのがわかる．発話時の Up/Dp パワー比曲線は特異値が多く，そのまま比較して見ただけでは有声音部の F0 との相関が読み取りにくい．

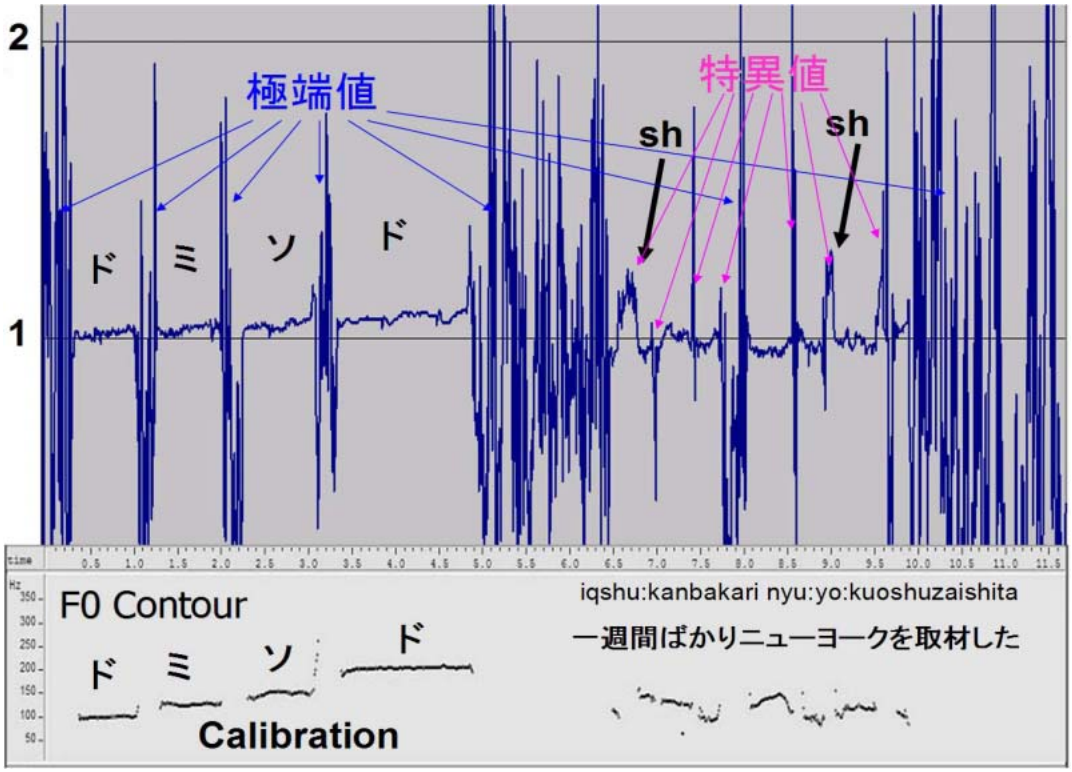


図 4.10　BTOS の Up/Dp パワー比曲線と F0 曲線との対比

そこで図 4.11 に単独サンプルの F0 値をもとに Up/Dp パワー比曲線の有声音部区間だけを取り出し，F0 と Up/Dp パワー比の散布図を描き，相関係数を求めてみた．相関係数は 0.683 であり高い相関を示した．

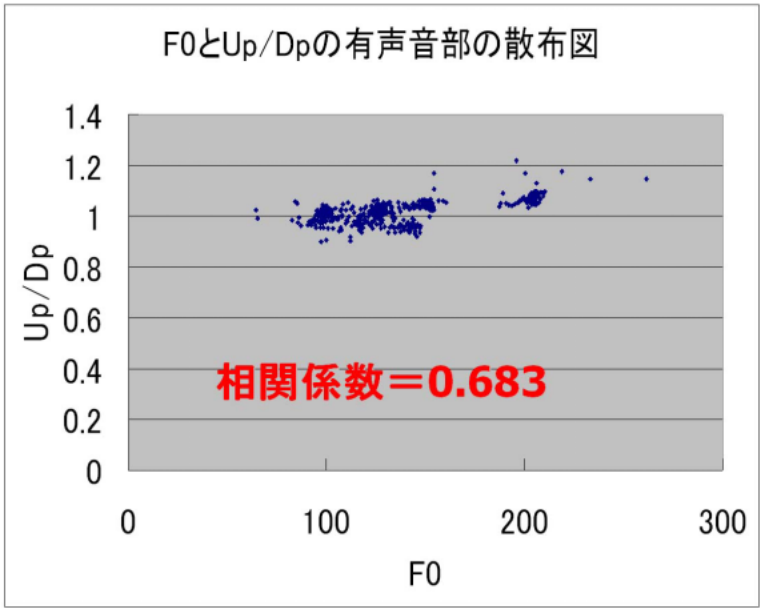


図 4.11　BTOS における F0 と Up/Dp パワー比の相関

BTOS については SOL 法による Up/Dp 曲線は有声音部において F0 と相関の高いパラメータであり，どれくらい高さの声を出そうとしているかを表現する別の指標となりうると言える．

4.5.2　NAM の Up/Dp パワー比

BTOS では SOL 法の基本的なアイデアが功を奏したが, NAM ではどうであろう．音源が声門裂に集中している BTOS と比して，その音源である乱流雑音の最強点が曖昧になることは当然想像される．のみならずパワーそのものが小さく S/N 比が低いため，誤差も多くなるであろう．また NAM と通常

音声は同時録音ができないため，基準となる F0 が得られない．別に収録した必ずしも同じ発話でない通常音声の F0 と比較せねばならない．

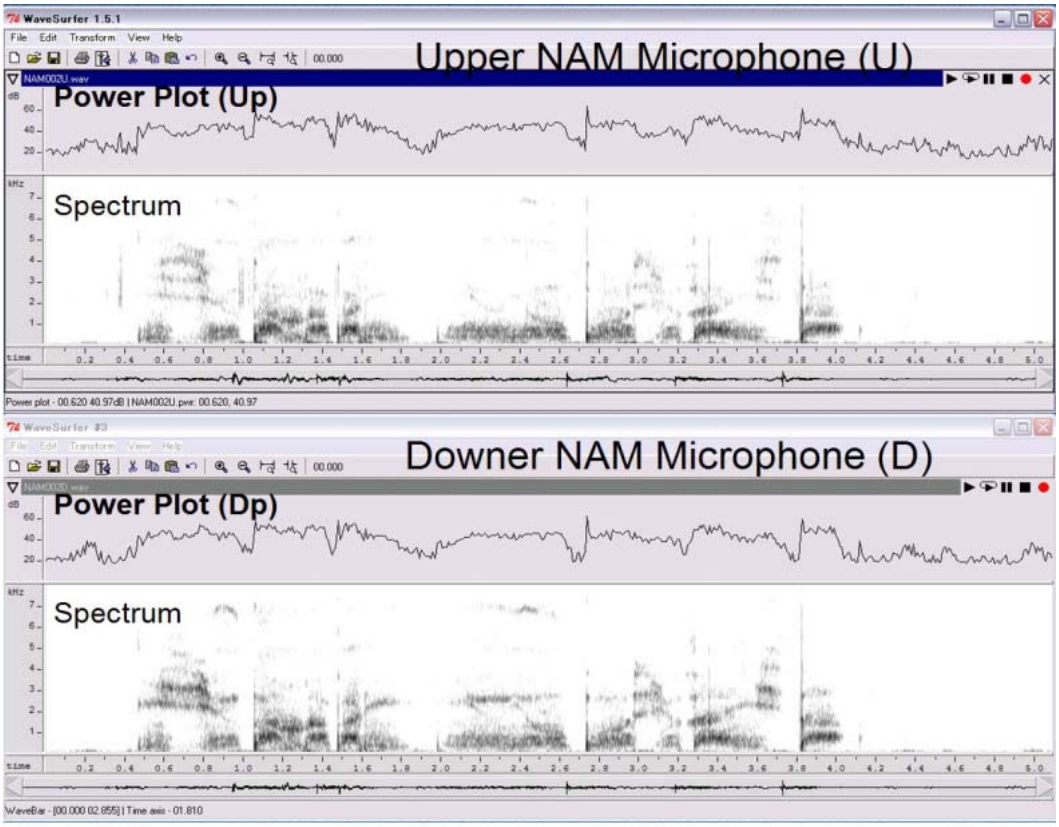


図 4.12　縦アレイ NAM マイクロフォンでステレオ収録した上下 NAM 音

先程の BTOS と同内容の文章「一週間ばかりニューヨークを取材した」を NAM 発話で読み上げ，縦アレイ NAM マイクロフォンで上下ステレオ同期収録を行った．BTOS 音と NAM 音ではパワーがかなり異なるため，NAM 音サンプリングに適した出力レベルに調節した. 図 4.12 は上下 NAM マイクロフォン（U と D）で収録した NAM 音のスペクトラムとパワーである．

図 4.13 は，NAM の Up/Dp パワー比曲線である．上部 NAM マイクロフォンから得られた NAM 音のスペクトラム，パワー，F0 プロットと同期させ

て表示した．当然のことながら NAM では F0 曲線は描出不能である．

図 4.14 に先程の BTOS での同内容発話の Up/Dp パワー比曲線を比較のために再掲する．NAM と BTOS で，同内容発話時に極端値，特異値も含めてほぼ同じパターンの曲線が描けているのがわかる．

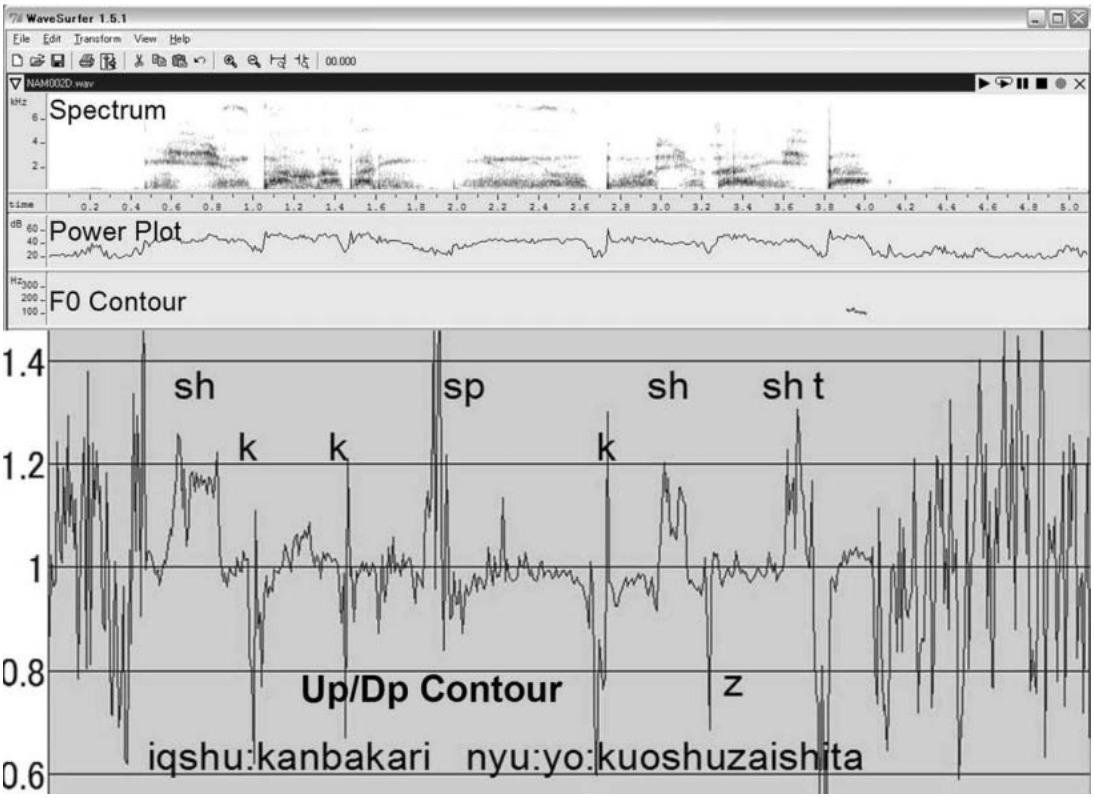


図 4.13　NAM の Up/Dp パワー比曲線

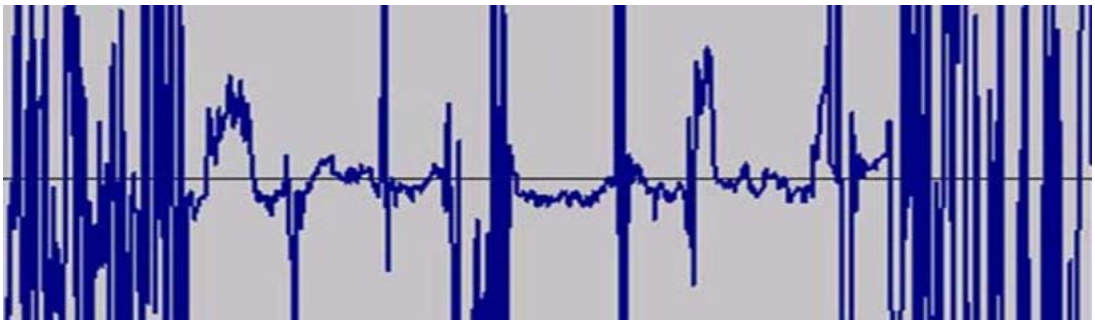


図 4.14　同内容発話の BTOS の Up/Dp パワー比曲線

NAM の Up/Dp パワー比曲線において無音部の極端値や無声子音の特異値は BTOS のそれと同様であるが，NAM では「z」のような有声子音にあたる音素でも特異値をとっている．「sh」や「k」等の無声子音はそれぞれ再現性をもって特異値をとっていることがわかる．これらの有声・無声子音による特異値や，ポーズや促音による極端値が，従来の「F0 のとぎれ」の部分に相当し，これらをのぞいた Up / Dp パワー比曲線が，喉頭部上下動の相対的位置，すなわち NAM 発話時のピッチ準備状態，韻律意図，もしくは同内容を BTOS 発話したときの喉頭上下動の「クセ」を表現しているものと考えられ，これを元にして NAM 発話のピッチ予測が可能ではないかと思われる．

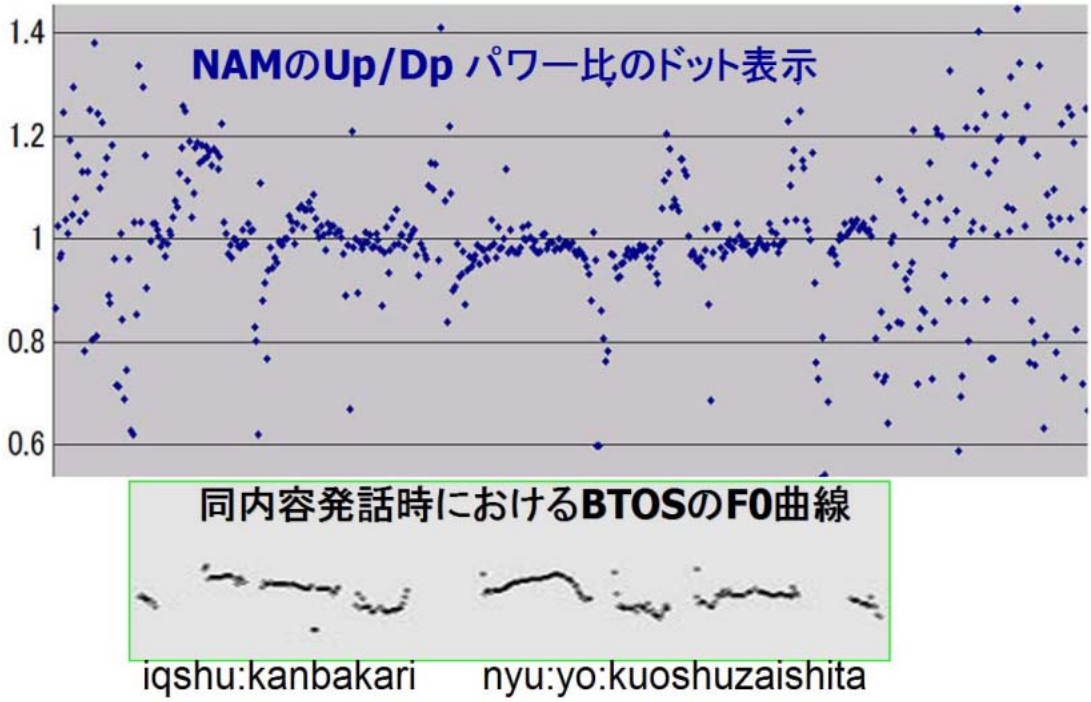


図 4.15　NAM の Up/Dp パワー比のドット表示と BTOS の F0 との比較

図 4.15 にこの Up/Dp パワー比をドットでプロットしたものも掲げるが，有声音に当たる部分はドットが集約しているため，擬似 F0 曲線として見やすいものとなる．図 4.15 の下段の BTOS での同内容発話の F0 曲線の形状と

比較して，この例ではその相似を視認できる．ただし NAM と BTOS は同時発話ができないため，同内容発話とはいっても時間伸縮の問題があり，意識的に同じスピードで読み上げてはいるが，単純に相関を数値的に表すことは難しい．

通常音声での読み上げと NAM での読み上げは，話速を一致させることは不可能であり，また単独 NAM 音情報からのでのピッチ情報がなく，また本当に通常音声発話時と同じ韻律，同じ喉頭の上下動で読み上げているかを判別できないため，ここでは単純に相関をとることをやめ，同じ音素バランス文 4 個ずつを，二人の男性が NAM と通常音声で読み上げたものを用いて，通常音声の F0 と NAM の Up/Dp をプロット表示で比較した例を掲げるに止めることにする．たまたまうまく行った例で数値的相関をひねり出すよりも，できるだけ多くの例を見てもらう方が，この手法の将来のために大切であると考えたからである．

図 4.16 が F0 と Up/Dp プロットの計 8 個の比較であるが，上は通常音声の F0，下が Up/Dp である．話者は NAM 発声のベテランである筆者 N（左列）と NAM 発話収録経験のない男性話者 H（右列）である．なお NAM 発話収録内容をリアルタイムでヘッドフォン聴取せず，発声に明確なフィードバックがかからぬようにした．その方がむしろ通常音声発話時の喉頭部上下動のクセが反映されると考えたからである．NAM マイクロフォンは上を最適装着位置に固定し，マイク間隔は二人ともマイク中心の間隔で 3cm，同じ NAM マイクロフォンセットを用いて実験を行った．比較的通常音声との形が似ていると感じられる Up/Dp プロットもあるが，そうでないものもある．

極端値や特異値は収録の条件によって大きく異なるが，無音部が極端値，子音相当部は極端値の原則はほぼ再現されている．むしろ有声音部のプロットが散らばって特異値に見えてしまう部分もある．有声音部にあまりばらつきが多いと視認では曲線としての上下動がわからなくなる．いずれにせよ，NAM のピッチ予測が，SOL 法で可能かということに関して，今の段階で結論めいたことは言えない．

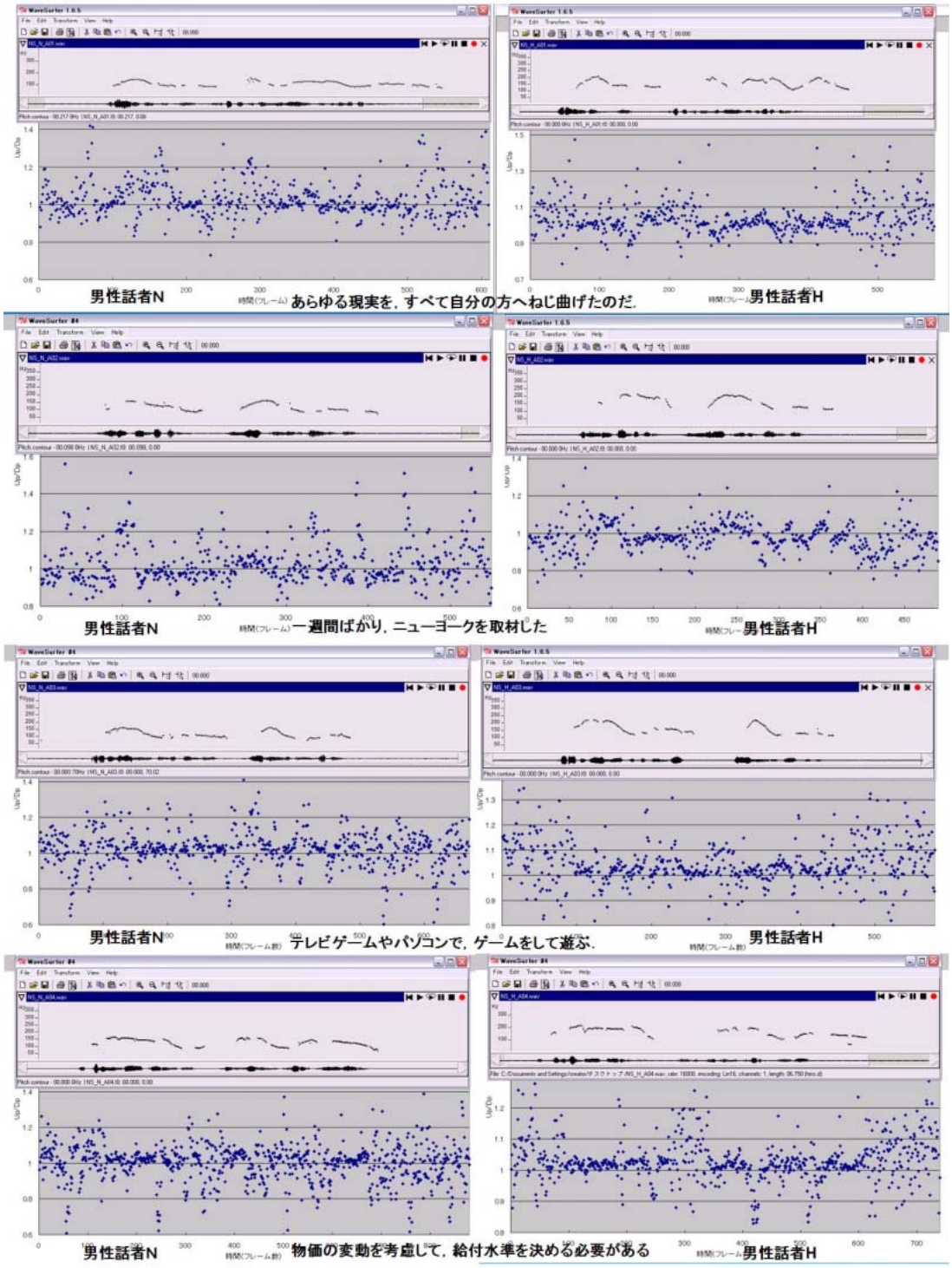


図 4.16　男性二話者による通常音声の F0 と Up/Dp 比

4.5 まとめと考察

同側の縦アレイマイクロフォンにより収録された二つの肉伝導音声でピッチ変化に伴って上下動する喉頭の位置を移動音源定位する SOL 法で，ピッチ予測を試みた．ここでは上下サンプルのパワー比を用いている．

BTOS 発声においては SOL 法による Up/Dp 曲線は有声音部において F0 と相関の高いパラメータであり，どれくらい高さの声を出そうとしているかを表現する別の指標となりうる．

NAM については，現段階で確固たることは言えないが，視覚的に有声音相当部を目で比較したときに，その曲線の上下動に似ている部分が多いと言える場合がある．曲線から極端値や特異値をカットし，スムージングや正規化を適切なアルゴリズムで行って，それを NAM の通常音声化の過程で F0 に変わるパラメータとして用いたときの，出力音声の自然性と判別性が最も良い評価の指標となろう．

現段階では未だに上下 NAM マイクロフォンの装着場所とマイク間距離を決めかねている段階であり，その位置と距離が最適かはまだ不明である．

縦アレイ NAM マイクロフォンは 3cm の距離であれば小型の単独デバイス内に組み込むことが可能であるので，今後種々の検討を行いたい．

NAM や BTOS を縦アレイ NAM マイクロフォンで常にステレオで収録しておくことによって, F0 とは違った次元の韻律情報を提供できるだけでなく, 特異値などから子母音の境界や子音の発生部位などの発声に関する種々の体内情報を推定することも考えられる．また体内雑音や遠方からの外部雑音の特定分離など，NAM マイクロフォンをアレイ化することの意義や応用範囲は非常に大きいと考える．人間を音声の発生源としての点と捉えるのではなく，大きな肉媒体の音場空間として捉えて体表から音を様々な角度からサンプリングしたとき，そこには非常に興味深くて役に立つ，深い世界が広がっていると思う．そしてマイクロフォンアレイ，音場の技術蓄積を肉伝導音声の世界に生かす場は限りなくある．

第5章 結語

5.1 まとめ

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，調音呼気音の肉伝導を「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ーフェースとする可能性を提示した.

またソフトシリコーン伝導型 NAM マイクロフォンというソフトシリコーンという人間の肉と音響インピーダンス的に近いゲル素材を音媒体に用いた全く新しいセンサーを開発することで，NAM の音質を向上させ，その実現に向けて大きく一歩を踏み出した.

肉伝導音声の世界は想像以上に奥深い．NAM マイクロフォンをステレオにするだけで，音声の違った側面が見えてくる．それは新しいピッチ情報のパラメータの発見に止まらない．音声の発生源である人間を，単に音を放散する「点」としてとらえる見方の癖を転回することができれば，そこには大きな肉媒体の音場の世界が広がっている.

5.2 NAM Interface Communication の現況と未来

第一章の 1.9 で NAM Interface Communication について触れた．NAM, BTOS 認識の「無音声認識」「総合発話認識」の大きな流れ，NAM の通常音声化による「無音声電話」の大きな流れ，この二つの幹は，先端大の音情報

処理学講座が中心となって企業においても音声認識，音声合成を専門とする方々の手により，着実にその基礎固めと展開がなされつつある．NAM サンプルのみで作られた不特定話者モデルも作られる予定である．BTOS も含めた形，さらには合わせて小声やパワーの大きい NAM など，不特定発話状態モデルの作成も望まれる．GMM による声質変換を用いた通常音声化も音質の向上には発明者自身が驚くばかりである．コーバスベースの波形接続型の speech to speech の手法も残されているので，これらが競争で質の良い通常音声を作り上げて欲しい．NAM マイクロフォンの改良と無線化，日常用途の普及製品版の大量生産は，企業の役目である．すでに数社が参入しているが，お互い競い合いながら良いものを創って頂きたいものだ．

他にも概念図の右脇に英語で挙げたトピックももしかしたら，この二つの幹以上に発展性のあるテーマかも知れない．NAM マイクロフォンの高感度の電子聴診器としての特性を生かし，心拍や呼吸音などの人間にとってバイタルな生体情報をモニタリングするという試みも，すでにある大学と研究機関の共同で推し進められている．利点は検査機器を検査目的で体表装着するのではなく，コミュニケーション目的で（つまり一日の大部分）装着しているデバイスから本来用途は別に情報が入手できる点であり，手法を確立すれば大規模な研究が容易である．言語以外の意図的体表雑音によって肉伝導ならではの意味のある情報伝達を作り出す体伝導パラ言語表現の分野はまた手つかずの状態である．しかしいずれ誰かがやるであろうし，体表雑音による行動モニタリングも生体情報モニタリングと合わせて，人間の行動パターン分析や危機回避など応用範囲は情報科学以外の方々にも参加の道がある．

そして実は最も大切なのは発声障害をもつ方々への応用である．これは自分が大学院でのこの過程を終わったら，中心として三年間は取り組んでいきたい課題であり，沢山の障害者の方々からの手紙は，すべてに返事を書き，一つ一つ大切にとってある．

最後に警察やマスコミの方々（「スパイマイク」として取り上げられた）など情報漏洩を何よりも嫌う職場の方々に興味を持ってもらったのも当然のこ

とであるが，ひとつだけ願わくは軍事目的で使用して欲しくないということである．テロリスト達に使われるのも恐れる（使いたいと思うだろうが)．

今のところ NAM Interface Communication は数本の枝分かれをした若木にすぎない．まだまだ枝分かれをするポテンシャルを秘めていると私は思うが, 社会のインフラの状況とそれよりも「人の心のインフラ」が整わないと，すぐに枯れてしまう可能性もある．

しかし 21 世紀の初頭，ここにしっかり種は残した．物まねが多いと揶揄される日本人が残した新しいコミュニケーションの種である. 大切に育てて，いつか大樹となる日が来たら，その木陰で幹に寄り添い，吹くそよ風にかき消されるほど静かにどこかのだれかと語り合いたい．

あとがき雑感

テレパシーなどという話題の書き出しで始まった博士論文は，前代未聞であると思うが,「あとがき」なのでもう一度触れよう．骨伝導スピーカーを使用してみるとわかるが，音は頭蓋骨全体を伝導するため，左右の内耳に同程度に音は伝わる．その結果左右の聴覚で音の定位ができず，頭の中で直接音が響いているような不思議な感覚がある．NAM マイクロフォンと NAM の通常音声化技術，そして骨伝導スピーカーとを組み合わせれば，相手が口だけを動かして聞き取れない声が，直接自分の頭の中に響いてくるような，それこそテレパシーと呼ぶに近い体験をすることになると思う．私にはなんでも先走って考えるクセがあるが，きっと 23 世紀の人達はこの論文を発見して「当然のことを……何をいまさら」とひそやかに NAM してくれると思う．

この論文の三つの章立て，つまり第二章，第三章，第四章の目的は，音情報処理学の大きな三つの潮流とそのまま重なる．すなわち，音声認識，音声合成，音場・マイクロフォンアレイである．また言語科学講座の韻律や感情音声というテーマにもはずれていないと考える．なぜなら NAM は「まわりに気兼ねする」という大切な感情音声だからである．この二つの講座のどの専門分野の方も，この NAM を用いた技術に参画できる．

それぞれについて十分な定量的また統計学的検討を行う余裕も実力もなく，単に方向性ときっかけとなるアイデアを示したに過ぎない嫌いはあるが，それが私の役目である．NAM に関するこの三つの基本となる考え方は，この分野の素人ゆえの自分，情報科学と医学という二つの分野の狭間にいる自分でなければ見いだし得なかったという自負がある．ひとつひとつをじっくりと突き詰めていけば，それぞれが深みのあるテーマに発展させうると信じる．

自分はただ，仕事を持ちながらの短い四年間の内に，音情報のプロフェッショナル達が集うこの世界の中で，自分でなければできない仕事，自分が一番役に立てる仕事を精一杯やったつもりである．

この論文では自分の大学院生活の総括として NAM をインターフェースとして用いる新しいコミュニケーション法の発見の経緯と，その豊かな可能性を紹介したが，その締めくくりとして，是非とも伝えたいことがある．

NAM という音韻は古代サンスクリット語で「〜に帰依する」「〜にすべてを捧げる」という意味の言葉の語幹である．この音韻は今のインドでも「ナマステー」という挨拶の言葉の中に生きている．日本の仏教などでも「南無阿弥陀仏」「南無妙法蓮華経」「南無八幡大菩薩」などと「南無」という当て字を冒頭に付けるのは，これに由来しているが，昔から多くの人々に使われてきたれっきとした日本語化した外来語である．

口の中で，誰に聞かせるでもなくつぶやくという意味で，「なむ」という響きは日本人に語感として馴染みやすいと考え，自分は自分の見つけた概念にこう命名した．外来語ではなく日本語であるから漢字やアルファベットで書く必要はない．論文などに書く必要上，Non-Audible Murmur（NAM）と，偶然意味的にも語呂的にも座りの良い字を「当てている」だけである．

そして気付いている人もいると思うが，実は発明者の私自身が，二つの領域のはざまに位置する「インターフェース」そのものなのであった．サイエンスとしては「医学と工学とのインターフェース」，また具体的事物としては「人間と機械とのインターフェース」，そして時間的・歴史的には「コミュニケーションの旧時代と新時代とのインターフェース」. 自分自身がこれらのインターフェースそのものになって，この新しいコミュニケーションに，私は自分を捧げたい．この論文のタイトルを思い起こせば，すなわち曰く，**「南無・インターフェース・コミュニケーション」である．**

卒業すれば私は久しぶりに「普通のおじさん」に戻る．これはそのためのリハビリテーションとしての「おやじギャグ的洒落」であるからして，決して感心したり，社交的笑いを浮かべたり，固まったりしないように．

今後の方針としては，この技術そのものは，奈良先端大の音情報処理学教室やその他の研究施設，また多くの企業の方々に任せていれば，自然と枝分かれし，最後には枝に見事な花を咲かせてもらえると信じている．自分としては，声の障害を持つ方々への日常へのコミュニケーション用途への応用を現場でじかに患者さん達に接しながら工夫を凝らし，そこで得た知見を，幅広く健常者の日常用途に使えるシステムを作る際のノウハウとして，企業や研究機関にフィードバックしていきたい．要するに私は「自分が本当に心から使いたい入力インターフェース」を創っているだけなのである．

奈良先端大での学生生活は忙しく，つらいこともあったが楽しかった．元はと言えば仕事に倦み疲れ, ただただ学校時代がなつかしかっただけなのだ．黒板と先生を前にして教室でノートをひろげ，シャープペンシルを指でくるくる回しながら，まわりの級友達と授業を聞く．そんな生活がもう一度純粋にやりたかっただけなのである．何の野心も功名心もなかった．だから学生と同じように修士から入試を受けて入学した. プロ野球のドラフトではなく，２軍の入団テストを受けて, あこがれの野球選手になったようなものである．

ところが，ひょんなことから「因」を作ったら，まったく知らない分野の，数え切れないほど沢山の人々との不思議な思いもかけない「縁」が生まれた．これを「因縁」という．南無阿弥陀仏.

謝辞

音情報処理学講座の鹿野清宏教授には，主指導教官として日頃より丁寧にご指導頂き，私の研究環境を整え，NAM 研究を通しての企業や他研究機関との沢山のつながりを作ってくださいました．問い合わせが集中した時，マネジメントに近いことまでして頂いたことを心苦しく思うと同時に，深く感謝させて頂きます．何よりも誰よりも自分の研究を認めて頂いて，その発展の基盤を作って下さったのがうれしかったです．

画像処理学講座の横矢直和教授には，ご多忙の折，審査委員として貴重なご意見と，「興味深く見ている」との励ましの言葉を頂きました．3 階の喫煙所で NAM マイクロフォンと関係の大きい眼鏡型モニターの話を個人的にしていただいたことを懐かしく思い出します．厚く御礼申し上げます．

言語科学講座のニック・キャンベル教授にも多大なご指導とご助力を頂きました．研究テーマの関係上，後期課程の 2 年間はゼミナールなどで直接先生に結果を発表する機会が減りましたが，その度ごとに適切な助言とお褒めの言葉を頂いたことを感謝いたします．EUROSPEECH では質問時間に助けて頂いてありがとうございました．

同言語科学講座の柏岡秀紀助教授は，指導教官の中では一番気軽に何でも相談できる貴重な先輩として常に頼れる存在でした．ゼミで発表するときに柏岡先生に向かってしゃべることが多かったです．いろいろ工学分野の話では噴飯もののレベルの発言もあったでしょうが，寛大に新規性のある良いポイントを評価して頂ける姿勢は自分にとって, 何よりもありがたかったです．

鹿野研の皆様，本当にありがとうございました．猿渡先生，李先生，川波先生，年上の学生というのは，自分にも経験がありますが，扱いにくかったと思います．今から思うとつまらないことをいろいろ質問にお邪魔したと思

いますが，嫌な顔ひとつ見せずに丁寧に答えてくださいました．鹿野研ゼミで竹苗君や東君をボロ雑巾のように批評されたみなさんの意見は自分の研究を補強する上にも大きなアドバイスになりました．パニコスさんには聴診器型 NAM マイクロフォンのデータを極限まで生かして，緻密で大きな結果を残してもらいました．エフ・ハリスト（どうもありがとう）．

鹿野研でかつてこんなに顔と名前が一致した学年はなかった M2 の青年諸君，2 階の喫煙室でのざっくばらんな会話は楽しく，本当に参考になる情報が多かった．アメリカ追随の風潮にめげず，一緒に煙草文化を維持しよう．

何よりも一番お世話になったのは阿部さんかも知れません．不備な書類やなくした領収書など，母親よろしく尻ぬぐいをさせてしまいました．どうかいつまでも芳しい香りを周囲に振りまいてください．

最後に就職前の大事な 2 年間を，まだ海のものとも山のものともわからない NAM に関わり，ともに悩んでくれた竹苗くん，東くん，貝野さん，加藤くん，岩永くん，ありがとう．年齢的にも心情的にも君たちは後輩と言うより自分の子供のような感覚で接することができました．誕生日のサプライズは一生忘れません．いつの日にか「続プロジェクト X」の撮影の時には同窓会しましょう. 君らはどこに就職しようとも NAM 研究のパイオニア達です．

言語科学講座のみんな，ドクターの先輩でありながら，分野のズレからあんまり面倒見て上げられなくてごめんね．芦村さんその辺のところを全部芦村さんに押しつけて申し訳なく思っております．ありがとうございました．

学外で忘れてはならないのは庄境さん，藤巻さんをはじめとする旭化成の方々です．離れた場所でありながら，特許のことや製品開発の情報交換など何度も足を運んで頂きました．メールでの情報交換も大学院で得られるもの以上だったような気がします．一緒に苦労し新しいものを創っているという実感があります．これからもよろしくお願い申し上げます．

最後に全国から私に会いにわざわざ奈良先端大まで足を運んで下さった研究機関や企業のみなさん．そして励ましやご鞭撻のお便りを下さった，発声障害をもつ沢山の方々ありがとうございました．

参考文献

[1] P. Badin, P. Perrier, L. J. Boe, C. Abry: "Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences," The Journal of Acoustical Society of America, vol.87, no.3, pp.1290-1300, 1990.

[2] 張 志鵬, 真鍋 宏幸, 堀越 力, 杉村 利明: “HMM 及びケプストラム係数特徴による筋電信号を用いた無発声音声認識,” 信学技報, SP2003-104, pp.7-12, Oct. 2003.

[3] W. M. Eppler: "Realization of Prosodic Features in Whispered Speech," The Journal of Acoustical Society of America, vol.29, no.1, pp.104-106, 1957.

[4] H. Ferner 編: ”臨床応用局所解剖図譜　第 1 巻　頭部・頸部:,” 医学書院, 東京, 1966.

[5] 降旗 健治, 柳沢 武三郎: “唇部位に振動ピックアップを取り付けたときの音声伝達について,” 日本音響学会講演論文集, 2-2-4, pp.571-572, Jun. 1979.

[6] 降旗 健治, 柳沢 武三郎: “唇部位に振動ピックアップを取り付けたときの音声伝達について,” 電子通信学会, EA78-52, 1978.

[7] 後藤 真孝: “非言語情報を活用した音声インターフェース,” 情報処理学会研究報告 SLP-52, vol.2004, no.74, pp.41-46, 2004.

[8] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano: “Accurate Hidden Markov Models for Non- Audible Murur (NAM) Recognition Based on Iterative Supervised Adaptation,” Proc. ASRU, pp.171-185, 2003.

[9] P. Heracleous, Y. Nakajima, A. Lee, Hiroshi H. Saruwatari, K. Shikano: "Audible (Normal) Speech and Inaudible Murmur Recognition Using NAM Microphone," Proceedings of the 12th European Signal Processing Conference (EUSIPCO2004), pp.329-332, Sept. 2004.

[10] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano: "Non-Audible Murmur (NAM) Speech Recognition Using a Stethoscopic NAM Microphone," Proceedings of 8th International Conference on Spoken Language Processing (ICSLP2004), WeC2102p-6, pp.527-530, October 2004.

[11] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, H. Shikano: "Non-Audible Murmur (NAM) Recognition Exploiting Adaptation Techniques," 電子情報通信学会音声研究会信学技報, SP2003-170, pp.1-6, 2004-1.

[12] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano: "Recognition of Non-Audible-Murmur (NAM) based on iterative supervised adaptation," 日本音響学会講演論文集, 3-Q-3, pp.131-132,

Sep. 2003.

[13] P. Helacleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano: "Normal speech recognition exploting advantages of Non-Audible Murmur (NAM) microphone," 日本音響学会講演論文集, 3-Q-2, pp147-148, Mar. 2004.

[14] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano: "A baseline study on speaker- independent Non-Audible Murmur (NAM) speech recognition," 日本音響学会講演論文集, 1-P-8, pp.177--178, Sept. 2004.

[15] 東 剛生, 中島 淑貴, 川波 弘道, 猿渡 洋, 鹿野 清宏: ”非可聴つぶやき声(NAM)から通常音声への変換手法の検討, ”日本音響学会講演論文集, 3-2-6, pp.327-328, Sept. 2004.

[16] R. E. Hillman, E. Oesterie, L. L. Feth: "Characteristics of the glottal turbulent noise source," The Journal of Acoustical Society of America, vol.74, no.3, pp.691-694, 1989.

[17] 今井 信臣, 山崎 芳男: “圧電セラミック型心音センサの設計, ” 日本音響学会誌, vol.49, no.9, 1993.

[18] S. C. Jou, T. Schultz, A. Waibel: “Adaptation for Soft Whisper Recognition Using a Throat Microphone,” Proc. 8th International Conferences on Spoken Language Processing (ICSLP2004), WeC2102p.12, Oct. 2004.

[19] 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）,” 日本音響学会誌, vol.57, no.3, pp.210-214, 2001.

[20] 河原 達也, 住吉 貴志, 李 晃伸, 板野 秀樹, 武田 一哉, 三村 正人, 伊藤 克亘, 伊藤 彰則, 鹿野 清宏,: “連続音声認識コンソーシアム 2002 年度ソフトウェアの概要,” 情報処理学会研究報告, 2003-SLP-48-1, 2003.

[21] 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.

[22] G. M. Kuhn: "On the front cavity resonance and its possible role in speech perception," The Journal of Acoustical Society of America, vol.58, no.2, pp.428-433, 1975.

[23] 金井 孝幸, 粥川 大祐, 降旗 健治, 柳沢 武三郎: “骨導音の外耳道内音圧特性からみた頭部内伝搬,” 信学技報, EA2002-7,2002.

[24] 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,” 日本音響学会講演論文集, 3-1-10, pp.111-112, Oct. 2001.

[25] C. J. Leggetter and 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.

[26] 李 晃伸, 河原 達也, 武田 一哉, 鹿野 清宏: “ Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識, ” 電子情報通信学会論文誌, vol.J83-D-II, no.12, pp.2517-2525, 2000.

[27] A. Lee, T. Kawahara, K. Shikano: “Julius - An Open Source Real-Time Large Vocabulary Recognition Engine,” Proc. 7th European Conference on Speech Communication and Technology (EUROSPEECH2001), pp.1691-1694, 2001.

[28] 松田 勝敬, 森 大毅, 粕谷 英樹: “ ささやき母音のフォルマント構造, ” 日本音響学会誌, vol.56, no.7, pp.477-487.

[29] 翠 輝久, 駒谷 和範, 清田 陽司, 河原 達也, 木戸 冬子: “ 音声対話による大規模知識ベース検索システム－音声版ダイアログナビ－, ” 情報処理学会研究報告 SLP-52, vol.2004, no.74, pp.21-26, 2004.

[30] 真鍋 宏幸, 平岩 明, 杉村 利明: “ 筋電信号を用いた無発生音声認識, ” インタラクション 2002 論文集, pp.181-182, 2002.

[31] 真鍋 宏幸, 平岩 明, 杉村 利明: “ 無発声音声認識のための指輪型電極の提案, ” FIT（情報科学技術フォーラム）2002, K-27, pp.421-422, 2002.

[32] P. Monson and W.R. emlin: “Quantative study of whisper,” Folia Phoniatr. 36, Publisher, pp.53-65, 1984.

[33] P. Mokhtari, H. Pfitzinger, C. T. Ishi, N. Campbell: "Laryngeal voice

quality conversion by glottal waveshape PCA, ” 日本音響学会講演論文集, 2-P-6, pp.341-342, Mar. 2004.

[34] 中村 敬介, 西村 竜一, 李 晃伸, 猿渡 洋, 鹿野 清宏: “ 実環境音声情報案内システムにおける環境雑音および不要発話の識別, ”電子情報通信学会技術研究報告, SP2003-172, pp.13-18, 2004.

[35] 中島 淑貴: “ 口内行動 －発話器官の動態分析における超音波イメージングの有用性－, ” 音声研究 , vol.7, no.3, pp.55-66, 2003.

[36] 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, ” 電子情報通信学会論文誌, vol.87-D-II, no.9, pp.1757-1764, 2004.

[37] Y. Nakajima, H. Kashioka, K. Shikano, N. Campbell: “Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin,” Proc. ICASSP, pp.708-711, 2003.

[38] Y. Nakajima, H. Kashioka, K. Shikano, N. Campbell: “Non-Audible Murmur Recognition,” Proc. EUROSPEECH, pp.2601-2604, 2003.

[39] Y. Nakajima, P. Heracleous, S. Kiyohiro, N. Campbell: “NAM (Non-Audible Murmur) Interface Communication,” Special Workshop in Maui, 2004.

[40] 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," 情報処理学会研究報告 SLP-52, vol.2004, no.74, pp.33-40, Jul. 2004.

[41] 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, ” 日本音響学会講演論文集, 3-Q-12, pp.175-176, Mar. 2003.

[42] 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, Mar. 2004.

[43] 中島 淑貴, 柏岡 秀紀, ニック キャンベル, 鹿野 清宏: ”縦アレイ NAM マイクロフォンによる韻律表現, ”日本音響学会講演論文集, 3-1-5, pp.119-120, Sept. 2004.

[44] 永井 秀利, 中村 貞吾, 野村 浩郷: “ 自然言語インターフェースのための無発声音声認識への活用を目的とした表面筋電波形の分析, ”電子情報通信学会研究報告 TL2002-52, pp.25-32, 2003.

[45] 永井 秀利, 中村 貞吾, 野村 浩郷: “ 無発声ないし微発声音声認識のための表面筋電波形からのノイズ低減手法, ” 情報処理学会九州支部「火の国情報シンポジウム」, 2003.

[46] 永井 秀利, 竹下 舞, 中村 貞吾, 野村 浩郷: “ 無発声ないし微発声音声認識への活用を目的とした表面筋電波形の調査, ” 情報処理学会 65 回全国大会講演論文集:2F-7, 2003.

[47] 大須賀美恵子: ”心的状態の指標としての心拍・心拍変動, ” ヒューマンインターフェース学会誌, vol.6, no.1, pp.9-14, 2004.

[48] L. Rabiner and B.H. Juang: “FUNDAMENTALS OF SPEECH RECOGNITION,” PTR Prentice-Hall, New Jersey, 1993.

[49] 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.

[50] 鈴木 英男: “マイクロフォンを使うにあたって注意すべきこと,” 日本音響学会誌, vol.55, no.5, pp.377-381.

[52] 榊 正則, 佐脇 康之, 藤野 昭二, 渥美 智也: “骨伝導マイク及びスピーカーによる双方向同時会話装置の試作,” 日本音響学会講演論文集, 1-5-23, pp.513-514, Sep. 2002.

[53] 坂口 剛史, 細井 裕司: “呼気を要しないサイレント音声入力装置－実用化に向けての基礎的検討－,” 信学技報, SP2003-106, pp.17-19, Oct. 2003.

[54] 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.

[55] 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," 日本音響学会講演論文集, 1-1-16, pp.31-32, Sept. 2004.

[56] I. B. Thomas: "Perceived Pitch of Whispered Vowels," The Journal of the Acoustical Society of America, vol.46, no.2, pp.468-470, 1969.

[57] V. C. Tartter: "What's in a whisper," The Journal of Acoustical Society

of America, vol.86, no.5, pp.1678-1683, 1989.

[58] 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,” 情報処理学会研究報告 SLP-54, vol.2004, no.74, pp.157-162, 2004.

[59] 竹内 康人: “有声音を伴わない発音による会話音声の創出入力システム,” 信学技報, DSP2003-91, pp.13-18, Sep. 2003.

[60] 竹内 康人: “呼気の静圧も含むささやき声（無発声音声）の収録のための表面が平滑な風防に収容されたマイクロフォン,” 信学技報, DSP2003-92, pp.19-22, Sep.2003.

[61] M. Unser, and M. Stone.: “Automated detection of the tongue surface in sequences of ultrasound images,” J. Acoust. Soc. Am. (5), pp.3001-3007, May 1992.

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

[63] 山出 慎吾, 李 晃伸, 猿渡 洋, 鹿野 清宏: “雑音に頑健な音韻モデルと教師なし話者適応,” 電子情報通信学会技術研究報告, SP2002-124, pp.19-24, 2002.

[64] S. Yong, G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland: The HTK Book (for HTK Version 3.2.1), Cambridge University Engineering Department, 2002.

研究業績

学術論文

1. 中島 淑貴: “口内行動　－発話器官の動態分析における超音波イメージングの有用性－，” 音声研究 , vol.7, no.3, pp.55-66, 2003.

2. 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,” 電子情報通信学会論文誌, vol.87-D-II, no.9, pp.1757-1764, 2004.

国際会議

1. Yoshitaka Nakajima, Hideki Kashioka, Kiyohiro Shikano, Nick Campbell: “Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin,” Proc. ICASSP, pp.708-711, 2003.

2. Yoshitaka Nakajima, Hideki Kashioka, Kiyohiro Shikano, Nick Campbell: “Non-Audible Murmur Recognition,” Proc. EUROSPEECH, pp.2601-2604, 2003.

3. Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano: “Accurate Hidden Markov Models for Non- Audible Murur (NAM) Recognition Based on Iterative Supervised Adaptation,” Proc. ASRU, pp.171-185, 2003.

4. Yoshitaka Nakajima, Panikos Heracleous, Shikano Kiyohiro, Nick Campbell: “NAM (Non-Audible Murmur) Interface Communication,” Special Workshop in Maui, 2004.

5. Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano: "Audible (Normal) Speech and Inaudible Murmur Recognition Using NAM Microphone," Proceedings of the 12th European Signal Processing Conference (EUSIPCO2004), pp.329-332, Sept. 2004.

6. Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano: "Non-Audible Murmur (NAM) Speech Recognition Using a Stethoscopic NAM Microphone," Proceedings of 8th International Conference on Spoken Language Processing (ICSLP2004), WeC2102p-6, pp.527-530, October 2004.

研究会

1. 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," 情報処理学会研究報告, vol.2004, No.74 pp.33-40, Jul. 2004.

2. Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee,Hiroshi Saruwatari, Kiyohiro Shikano: "Non-Audible Murmur (NAM) Recognition Exploiting Adaptation Techniques," 電子情報通信学会音声研究会信学技報, SP2003-170, pp.1-6, 2004-1.

大会発表

1. 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,” 日本音響学会講演論文集, 3-Q-12, pp.175-176, Mar. 2003.

2. Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano: "Recognition of Non-Audible-Murmur (NAM) based on iterative supervised adaptation," 日本音響学会講演論文集, 3-Q-3, pp.131-132, Sep. 2003.

3. 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, Mar. 2004.

4. Panikos Helacleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano: “Normal speech recognitionexploting advantages of Non-Audible Murmur (NAM) microphone,” 日本音響学会講演論文集, 3-Q-2, pp147-148, Mar. 2004.

5. Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari , Kiyohiro Shikano: "A baseline study on speaker-independent Non-Audible Murmur (NAM) speech recognition," 日本音響学会講演論文集, 1-P-8, pp.177--178, Sept. 2004.

6. 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," 日本音響学会講演論文集, 1-1-16, pp.31-32, Sept. 2004.

7. 中島 淑貴, 柏岡 秀紀, ニック キャンベル, 鹿野 清宏: "縦アレイ NAM マイクロフォンによる韻律表現," 日本音響学会講演論文集, 3-1-5, pp.119-120, Sept. 2004.

8. 東 剛生, 中島 淑貴, 川波 弘道, 猿渡 洋, 鹿野 清宏: "非可聴つぶやき声(NAM)から通常音声への変換手法の検討," 日本音響学会講演論文集, 3-2-6, pp.327-328, Sept. 2004.

講演

1. Yoshitaka Nakajima, "Non-Audible Murmur Recognition (The New Universal and Individual Design Interface)," 2003 NAIST COE International Symposium, March, 2003.

2. 中島 淑貴, “無音声認識（ Non-Audible Murmur Recognition ）”, 2003 年 12 月 1 日, デンソー基礎研究所（愛知県豊田市）.

3. 中島 淑貴, “NAM という新しい人間の情報伝達手段 ”, 2004 年 3 月 8 日,旭化成情報科学研究所（神奈川県厚木市）.

特許

1. 国内特許出願番号：特願 2002-252421

2. 国際出願番号： PCT/JP03/11157

3. 国内特許出願番号：特願 2004-004163

4. 国内特許出願番号：特願 2004-270262

受賞

1. 2003 年 3 月 3 日　第 12 回 ISID 学生論文 IT “ 夢 ” 大賞　一位受賞.

2. 2003 年　日本音響学会春季研究発表会　ポスター賞受賞.

展示会

1. “ 音声対話認識，無音声認識，” イノベーション・ジャパン 2004, 2004 年 9 月 28 日〜30 日，東京国際フォーラム，C-24.

2. “ 音声コミュニケーションインターフェース，” CEATEC JAPAN 2004, 2004 年 10 月 5 日〜9 日, 幕張メッセ, ミツミ電機ブース. ミツミ製 NAM マイクロフォン試作品展示.

3. “ 音声コミュニケーションインターフェース，” 第 11 回 ITS 世界会議 愛知・名古屋 2004, 2004 年 10 月 18 日〜24 日, ポートメッセ名古屋, NTT DoCoMo ブース. ミツミ製 NAM マイクロフォン試作品展示.

マスコミ・メディア

1. 2004 年 4 月 5 日,「つぶやき声もキャッチ」, 読売新聞夕刊.

2. 2004 年 4 月 27 日放送, RADIO JAPAN 火曜フォーカス(テクノロジー), NHK 国際ラジオ放送.

3. 2004 年 5 月 10 日放送, 毎日放送「ちちんぷいぷい」.

4. 2004 年 6 月 4 日掲載,「音声認識の新たな方法」, 奈良新聞朝刊.

5. 2004 年 6 月 10 日掲載,「ひと往来」, 奈良新聞朝刊.

6. 2004 年 6 月 19 日放送, NHK ラジオ第一放送「カルチャー＆サイエンス」.

7. 2004 年 7 月 31 日掲載,『“ 無音声入力マイク ” を開発（ひと人抄)」, 読売新聞夕刊.

8. 2004 年 10 月 8 日放送, 奈良テレビ「ざっくばらん」.

講習会（NAM マイクロフォン製作研修）

1. 2004 年 4 月　カーネギー・メロン大学：Szu-Chen Stan Jou

2. 2004 年 5 月　旭化成：庄境誠　藤巻栄

3. 2004 年 6 月　国際電気通信基礎技術研究所（ATR)：小作浩美　野間春生　納屋太　小暮潔

企業・研究機関へのデモ・プレゼン・データ提供等

（自分自身が直接応対したもののみ，大学院広報や知財部への問い合わせや音情報処理学講座のみの応対等は省く）

2003年　旭化成情報科学研究所，電通国際情報サービス，カーネギー・メロン大学，Panasonic ヒューマンウェア研究所，Panasonic 音響グループ，日本 IBM，TIS，奈良県企画部学研協力課，南都銀行，豊田中央研究所，東芝，NEC，NTT，デンソー，ホシデン，きんでん，MITSUBISHI，関西経済連合会

2004年　ATR，CASIO，NHK，KDDI，ノキア・ジャパンリサーチセンター，NTT DoCoMo ユビキタスサービス部，ミツミ電機，豊橋技術大学情報工学系，サン・マイクロシステムズ，NEC ラボラトリーズマルチメディア研究所，大阪工大情報メディア学科，奈良県立医大耳鼻咽喉科学教室科，日産自動車総合研究所，YAMAHA，名古屋工大，言語聴覚士協会，銀鈴会，奈良テレビ，読売新聞，奈良新聞，東北大学歯学研究科

2005年　NTTDoCoMo 移動機開発部（旭化成との共同プレゼンテーション）