

# 無音声認識 (NAM 認識) におけるセンシング方法の改善\*

◎中島淑貴 (奈良先端大 情報) △柏岡秀紀 (奈良先端大/ATR)  
 ニックキャンベル (奈良先端大/ATR/CREST) 鹿野清宏 (奈良先端大 情報)

## 1 はじめに

我々は非可聴の調音呼吸音である非可聴つぶやき (Non-Audible Murmur: NAM) を定義し、これを耳介後下部の皮膚に接した NAM マイクロフォンよりサンプリングして HMM による大語彙連続認識が可能であることを示した。

NAM は認識だけではなく、モーフィングや音源付与などによる通常音声化が可能であり、無音声携帯電話などへの応用が期待されている。そのためには従来の聴診器型 NAM マイクロフォンでサンプリングされる NAM は 2KHz 以上にフォルマントがほとんど見られず、その帯域の広範囲化と感度の向上などの音質向上が不可欠である。今回我々は NAM マイクロフォンの抜本的な改良を行い、その画期的な音質向上と、さらなる外部雑音への頑強性を得たので報告する。\*

## 2 聴診器型 NAM マイクロフォン

図 1 が従来の聴診器型 NAM マイクロフォンである。コンデンサマイクロフォン (以下 ECM) を用いており、体表に接着した振動板と ECM の電極との間に聴診器と同じ原理の微小密閉反響空間が存在する。このセンサーでは第一、第二フォルマントまでは描出され、連結学習や Iterative MLLR による話者適応で、外部雑音に頑強な実用的な大語彙連続認識は可能であるが、そのままの形で人間の耳には聞き取りにくい欠点があった。これを用いて 16KHz サンプリングした NAM と BTOS (体内伝導通常音声) のスペクトラムが図 7 と図 8 の B である。対照のため通常 ECM で採取した空気伝導の NAM と通常音声で A に掲げる。

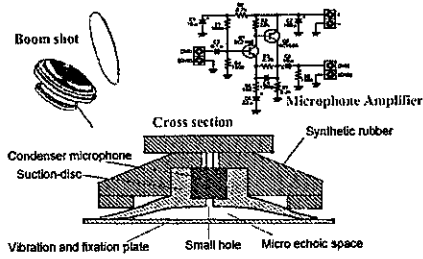


図1 聴診器型 NAM マイクロフォン

## 3 新 NAM マイクロフォンの工夫

聴診器型の場合、ECM と皮膚表面までにポリエステル製振動板と陰圧の空気が介在しており、これによる帯域制限の可能性が考えられた。そこで形成が容易で医療用にも多用されるシリコンのみを ECM と皮膚との間に介在させる実験を試みた。ただし市販の ECM は表面の小孔を通じて振動電

極を空気が振動させる構造となっているため、前面の金属を切断し、振動電極が表面に露出する構造とした。これを仮にオープンコンデンサーマイクロフォン (OECM) と呼ぶ。また ECM は前面からだけではなく、背部や周囲からも振動を拾うため、シリコンで OECM 全体を包み込んだ。

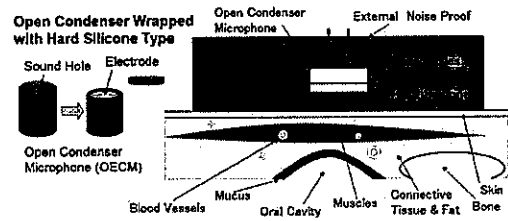


図2 ハードシリコン包埋オープンコンデンサー型 (OCWHS)

これを聴診器型と同様に耳介後下部の最適 NAM サンプリング位置に密着させ (接着はしない)、採取した NAM と BTOS が図 7,8 の C である。感度は聴診器型よりやや落ちるものの 6KHz 周辺までスペクトラムを淡く描出可能となった。

## 4 軟部組織の音響インピーダンスとシリコン

人間の軟部組織、いわゆる肉の音響インピーダンスは水に近い。音響インピーダンスの異なる物質同士の界面では音の反射が起こるため、シリコンの種類や硬さも音響インピーダンスをできるだけ軟部組織に近づければ反射による減衰を最小限にすることができる。

様々な種類と硬さのシリコン円盤 (厚さ 5mm) を用意し、医療用超音波イメージング装置を用いて、プローブと皮膚の間に図 3 の如くシリコン円盤をはさんで、音響インピーダンスの差を可視化してみた。円盤を通して背部の体内構造が描出できれば、音響インピーダンスは軟部組織とほぼ同等であることがわかる。前掲のシリコンでは後方に音響陰影ができ、体内構造は淡くしか描出できないのに対し (図 3 左)、人肌の柔らかさのソフトシリコンでは後方の体内構造がほぼ鮮明に描出された。これを用いれば、人体に OECM を埋め込んだ場合と音響的にはほぼ同様の効果が得られると考えた。

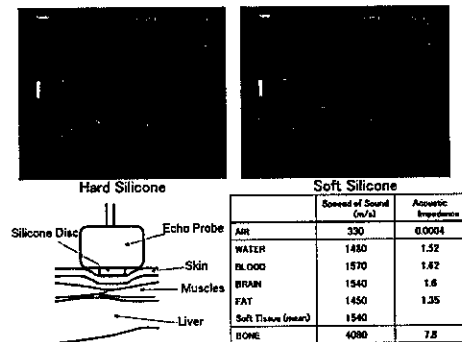


図3 超音波診断装置で見る人体とシリコンの音響インピーダンス

\* "Remodeling of the Sensing Methods of NAM Recognition"

by Yoshitaka Nakajima (NAIST), Hideki Kashioka (ATR/NAIST), Kiyohiro Shikano (NAIST), Nick Campbell (ATR/CREST/NAIST)

## 5 ソフトシリコンを用いたNAM マイクロフォン

図4, 5, 6 が、前掲の人肌の柔らかさと弾性に近いソフトシリコンを用いた NAM マイクロフォンである。円盤状の圧電素子を用いたもの(図4)や、聴診器構造の微小密閉反響空間をソフトシリコンに置き換えたもの(図5)、OECM をソフトシリコンで完全に包み込んだもの(図6)など様々なタイプを試作した。

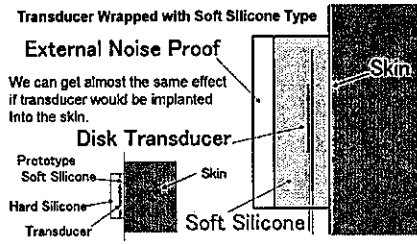


図4 ソフトシリコン包埋圧電素子型(TWSS)

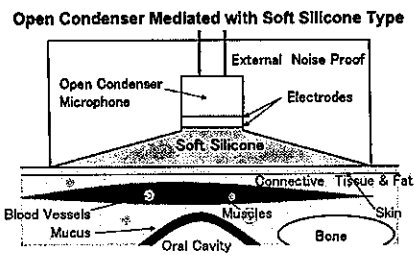


図5 ソフトシリコン伝導オープンコンデンサー型 (OCMSS)

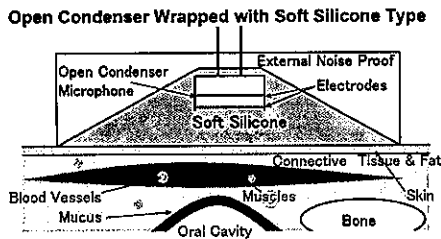


図6 ソフトシリコン包埋オープンコンデンサー型 (OCWSS)

## 6 NAM マイクロフォン帯域の広範化

図7, 8 は以上の NAM マイクロフォンで 16KHz サンプリングされた NAM と BTOS のスペクトラムである (A は対照)。

図3~6 のどの型の新 NAM マイクロフォンも従来の聴診器型(B)より帯域が広範化し、より高いフォルマントが現れた。

図7, 8 の D に見られる様に、圧電素子を用いたものは帯域が非常に広範であるが、皮膚にかなり強く圧着せねばならず、安定性が OECM を用いたものより劣った。感度と安定性において最も優れたものは図6の如く OECM をソフトシリコンで包埋したタイプのものであった(図7, 8 の F)。

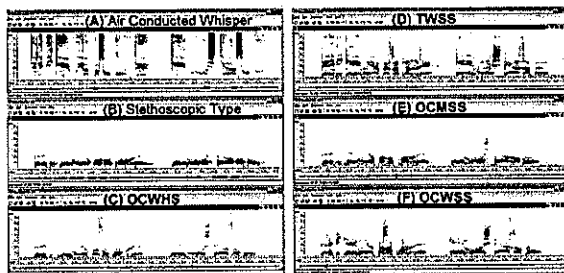


図7 新NAM マイクロフォンによる NAM スペクトラム (C, D, E, F)

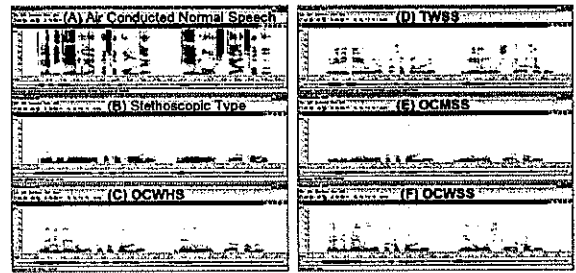


図8 新NAM マイクロフォンによる BTOS スペクトラム(C, D, E, F)

## 9 外部雑音に対する頑強性

外部雑音として大音量の繰り返し TSP 信号を用い、NAM マイクロフォンの外部雑音に対する頑強性を評価した。TSP 信号を用いているが、いわゆるマイクロフォン特性とは異なる。すべて同じマイクアンプを用い、認識の可能な NAM がサンプリングできるゲインに調節し、NAM も同時収録した。新 NAM マイクロフォンが、従来の聴診器型に比し特に 1 KHz 付近の第一フォルマントに近い部分での雑音に対し頑強であることがわかる。A は対照としての通常マイク音である。

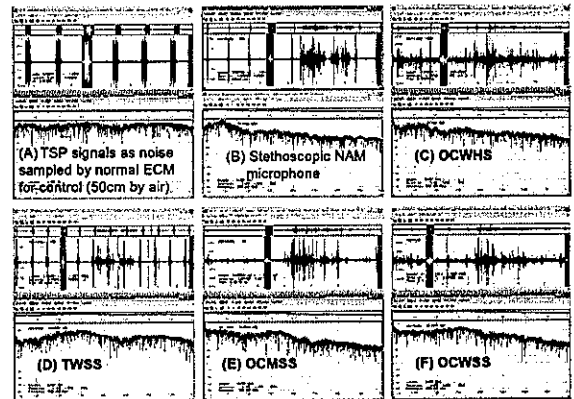


図9 NAM マイクロフォンの外部雑音に対する頑強性

## 10 まとめ

これらの新 NAM マイクロフォンでサンプリングした NAM や BTOS は、そのままの形で人が聴取してもほぼ完全に聞き分け可能である。認識だけでなく、無音声携帯電話などを初めとする通信への広範な応用が可能と考えられる。

\*謝辞: 本研究を援助頂いた、JST/CREST と素材提供の歯科技工ラボラトリーピノキオ社長の徳田均氏に感謝します。

## 【参考文献】

- [1] 中島淑貴, 相岡秀紀, 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", 日本音響学会論文論文集, pp.175-176, Mar. 2003.
- [2] Yoshitaka Nakajima, Hideki Kashioka, Kiyohiro Shikano, Nick Campbell "Non-Audible Murmur Recognition", EUROSPEECH, 2003.
- [3] Panikos Heracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano, "Accurate Hidden Markov Models for Non-Audible Murmur (NAM) Recognition Based on Iterative Supervised Adaptation", ASRU 2003