# 特集　音声認識用骨伝導マイクロホンの開発*1

## Development of the Bone Conduction Microphone for Voice Recognition

山田芳靖　　土方啓暢
Yoshiyasu YAMADA　Yoshimasa HIJIKATA

A compact, high sensitivity acceleration sensor, which can realize voice recognition by bone-conducted voice, has been developed. The sensitivity from 2 through 4kHz was enhanced to overcome the low output of bone-conducted voice in this frequency range, thereby a high voice recognition ratio of over 80% has been achieved. In order to actualize such characteristics of the frequency response of the sensor, multi piezoelectric bimorphs with different resonant frequencies are utilized. LPC cepstrum distance (CD) between bone-conducted voice and air-conducted voice was calculated to evaluate the suitability with regard to voice recognition. Compared to CDs of conventional bone conduction microphones, CDs of the latest bone conduction microphones decreased, especially in Japanese vowels /i/ and /e/, plosive consonants such as /k/ and /t/, and spirant consonants such as /s/ and /h/. This result indicates that the bone-conducted voice from the microphones in existence at present has become closer to that of air-conducted voice.

**Key words** : Bone-conducted voice, Voice recognition, Noisy environment, Piezoelectric bimorph, Acceleration sensor

## 1．はじめに

車室内で増加している情報機器を安全に操作する手段として，音声認識技術への関心が高まっている．現在その認識率は，静寂環境では90%以上の数値が得られるようになってきているが，周囲の騒音で認識率が著しく影響を受けることが一般的に知られている．従来，SS（Spectral Subtraction）法など信号処理技術による対策はなされてきたが，充分な効果が出ているとは言えなかった．そこで今回我々は，原理的に周囲の騒音が入り難い骨伝導音を利用したマイクロホンによる対策を試みた．骨伝導音は，発話時に話者の頭骨や皮膚組織などを振動として伝搬して頭部表面で検出されるものであり，周囲の騒音が載りにくいという特徴を有する．現在，軍や消防士など特殊環境用としてや，一部イヤホンマイクとして市販されている．しかし，骨伝導音声による音声認識を実用化した報告は，これまでになかった．そこで，我々は，骨伝導音声が前述の問題解決に極めて有効な手段と考え，音声認識への適用可能性を検討した．

## 2．骨伝導マイクロホンの試作

### 2.1　骨伝導音とは

骨伝導音は前述のように，周囲の騒音に比べ話者本人の声をS/Nよく検出することができる．それは気体である空気と固体である頭部の間の音響インピーダンス（補足参照）の差による．従って，原理的に騒音下で用いるのには適しているが，今まで音声認識で実用化された例はない．それは音質が通常の音声（以後気導音）と比べ劣っているからに他ならない．

Fig. 1は，①気導音②骨伝導音③市販骨伝導マイク（NHC G-450）の出力の，周波数成分を示している．①はエレクトレットコンデンサマイク，②はフラットな特性を持つ工業測定用加速度センサ（電圧感度10mV/ms$^{-2}$）で測定しており骨伝導音そのものの特性である．これらは同一話者がATR音素バランス文*2を発話したときに測定した．この図によると，骨伝導音は気導音に比べ，1kHz以上の周波数帯域で約20dB出力が落ちていることが分かる．このためこもったような音になり音質は悪い．音声認識を行う上でも音声の特徴量が気導音とは異なったものになるため，認識率が落ちるという問題点があった．一方，現在市販されている骨伝導マイクロホンは，主に通話を行うことだけ目的としているため，Fig. 1に示すような周波数特性を示しており，2kHz以上の周波数帯域ではほとんど感度が無く，音声認識には全く適していない．そこで本研究の目的は，音声認識に適した骨伝導音を検出可能な骨伝導マイクロホンを実現することである．

*1 2003年3月20日　原稿受理
*2 音素バランス文
（1）すべての日本語発音記号が均一に入った文章　（2）音源（有声音，無声音）に対してすべての調音位置，調音方法で作られたあらゆる音声信号を取り出すための文章→周波数成分を正しく抽出するのに適している．

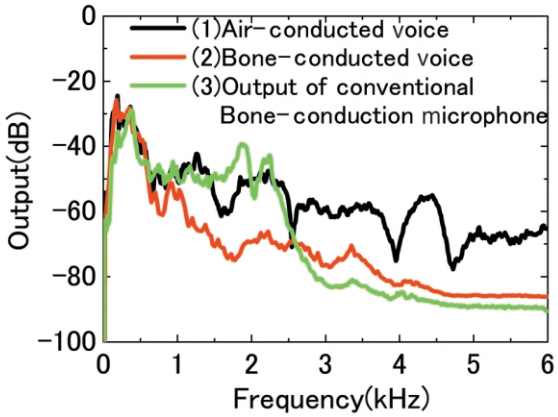


Fig. 1 Spectral component of (1) Air-conducted voice (2) Bone-conducted voice (3) Conventional bone conduction microphone

### 2.2　音声認識実現のためのマイクロホン仕様

定性的には，気導音に比べて出力が落ちている周波数帯域を増幅できる特性のマイクロホンが実現できればいいと考えられるが，どこまでの帯域が必要なのであろうか？母音を周波数解析すると，ホルマントと呼ばれる母音毎に特有の周波数が存在する．これは声帯から発せられた特定の周波数（ピッチ）とその高調波よりなる音声が口蓋内で共鳴することで，言葉によって特徴的に強調される周波数のことである．ホルマントは複数有り，低い周波数から第１，第２,,,,と名付けられるが，通常第２までで議論される．Fig. 2に日本語の母音毎の第１，第２ホルマントの分布を示す.[1] これによれば，第１，第２ホルマントは0.5〜3.5kHzの範囲に存在していることが分かる．そこで今回は，3.5kHzまでの周波数帯域で気導音並みの出力を実現する骨伝導マイクロホンを試作することとした．

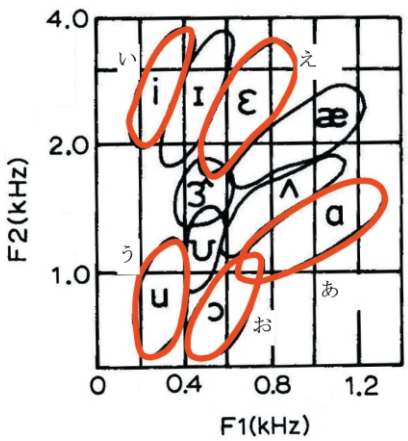


Fig. 2 Distribution of the first and second formant of Japanese vowels

Fig. 1の結果から，3.5kHz付近での骨伝導音（振動）の加速度レベルは約80dB（ $10^{-4}$m/s$^2$）である．この時の出力（電圧値）を約20dB（10倍）引き上げるとほぼ気導音と同じ周波数特性が得られると考えられるので，この周波数帯域において必要なセンサ感度は約100mV/ms$^{-2}$である．

振動を検知するセンサエレメントとして，市販骨伝導マイクロホンは圧電素子で金属板を間に挟んでサンドイッチ状にした圧電バイモルフを使用していた．コンパクトで高出力を得られるその特性は，今回の想定用途（車内外で，常時装着して音声入出力を行う）に合致すると考えられたので，今回も圧電バイモルフを使って試作を行うこととした．また装着方法としては，市販品等同様にイヤホンマイク様に外耳道に挿入して用いることにした．そのような条件下で圧電バイモルフを使ったときに得られる電圧感度の概念図をFig. 3に示す．通常センサ特性として良好なのは，周波数にかかわらずフラットな感度が得られる領域であるが，そのような領域では今回必要な感度が得られないことが明らかである．唯一共振周波数付近では，高い感度が得られるところから，今回はこの領域を使ってセンシングを行うことにした．ただし高感度が得られる帯域が狭いので，圧電バイモルフを複数本使うことでより広い帯域での感度補正を目指した．

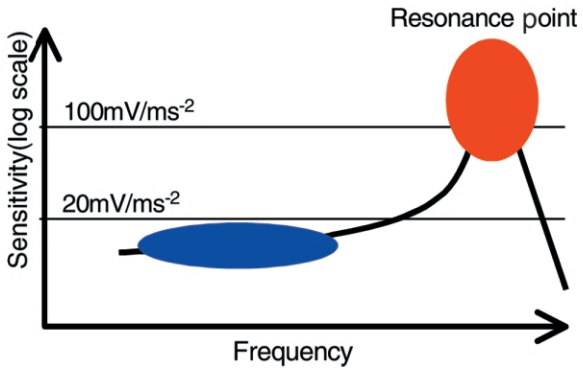


Fig. 3 Relation between frequency and sensitivity of piezoelectric bimorph sensor

### 2.3　試作の結果

今回，共振周波数の異なる複数の圧電バイモルフを実装するために，Fig. 4に示す新規構造を提案した．

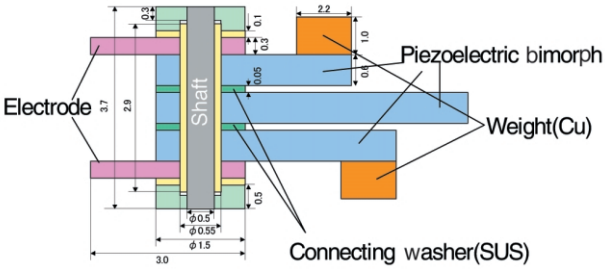


Fig. 4 Cross sectional diagram of a newly developed piezoelectric sensor structure

圧電バイモルフ同士は端面に穴を開けてシャフトを通して串刺しにして，両端からねじ止めして固定した．バイモフル間は導電性のワッシャを介することで，最小限の空隙を設けた上で電気的に接続している．この構造で，最大５本まで積層可能なことは確認した．圧電バイモルフの共振周波数，感度などは，基本的にその体格によって求まり，今回は有限要素法（FEM：Cybernet社のANSYSを使用）で設計を行った．電圧感度と周波数の関係の設計値をFig. 5に示す．

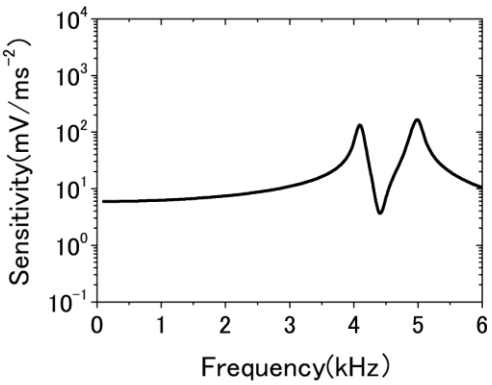


Fig. 5 Simulated frequency response of sensitivity of the suggested sensor

１本のバイモルフは4kHzに共振周波数を持ち，それより下の周波数に行くに従って，徐々に感度が低くなることで気導音に近い特性が得られるように狙った．また２本のバイモルフが5kHzに共振周波数を持ち，全体の感度を上げている．ピーク間の深い窪みは反共振特性によるものである．

Fig. 6に実際に製作した骨伝導マイクの写真（a）と，その電圧感度の測定値（b）を示す．電圧感度の測定値は，シミュレーションで得られた値と若干の誤差を含んでいるものの，およそ設計値通りの値が出ていると考えられる．

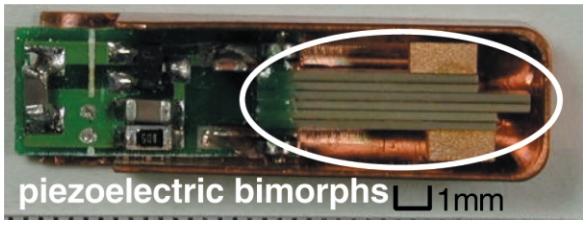


Fig. 6 (a) Photo of the developed multi bimorph sensor

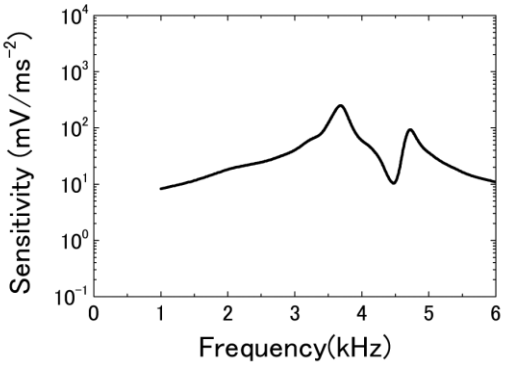


Fig. 6 (b) Measured frequency response of the sensor

## ３．音声認識実験

試作した骨伝導マイクロホンを用いて実際に骨伝導音を録音し，それを使って音声認識実験を行うことで効果の確認を行った．

### 3.1　実験方法

発声データの録音は，防音室内（暗騒音：23dB（A））において気導音と骨伝導音を同時録音することで行った．気導音は話者の口元から約30cmの地点にエレクトレットコンデンサーマイク（SONY ECM-T145）設置し，骨伝導音は市販品（NHC G-450）と今回の試作品をそれぞれ，外耳道内に挿入し外耳道の壁面から骨伝導音（振動）を検出した．気導音，骨伝導音それぞれをDAT（Digital Audio Tape）レコーダ（SONY TCD-D8）の左右チャンネルに入力して録音した．録音時のサンプリング周波数は48kHz，量子化ビット数は16ビットで，音声認識を行うときには12kHzまでダウンサンプリングした．発話した音声は，ATR音素バランス文10文と市販カーナビゲーションシステム用の音声認識コマンド100個である．音素バランス文は無発声区間を削除した残りの発声区間をFFT解析することにより，検出音声の周波数特性を調べた．また音声認識コマンドは市販のナビゲーションシステム（SONY NVX-M7000）のマイク端子から入力し，実際

に認識するかを実験し，100語のうち認識したコマンドの数で認識率を求めた．

### 3.2　実験結果

Fig. 7に今回試作した骨伝導マイクロホンで得られた骨伝導音の周波数特性を示す．

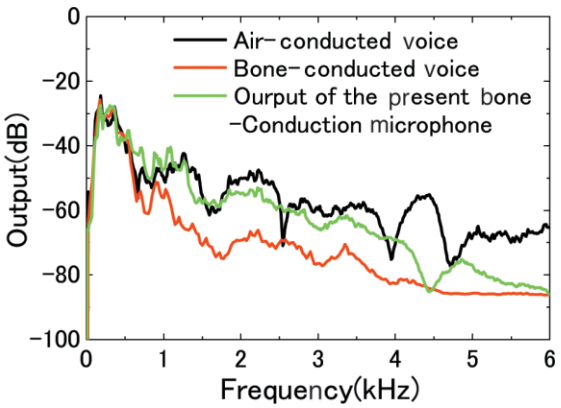


Fig. 7　Spectral component of the present conventional bone-conduction microphone

骨伝導音自体の周波数成分と比較して，1〜4kHzの周波数帯域において，気導音と非常に近い特性になっていることが分かる．実際に聞いた感じでも，骨伝導音特有のこもった感じが薄らいでいることが分かった．

このマイクロホンを用いて音声認識を行った結果をFig. 8に示す．話者A及びB（共に30代男性）で，市販の骨伝導マイクと今回の試作品を使って，防音室内で録音した音声で音声認識率を比較したところ，市販の骨伝導マイクロホンでは認識率40%前後であったが，試作マイクロホンでは80%以上の認識率が得られた．この数字は一般に音声認識率の許容レベルといわれる70%を越えている．

但し現在のところ，このような80%以上の骨伝導音による音声認識率が，すべての話者で得られているわけではない．人によってはまだ50%程度しか認識率が得られない場合もある．これについては次章で触れる．

## 4．考察

### 4.1　LPCケプストラム距離による評価

音声認識をする場合，入力された音声を短い時間で区切って音響分析し，そこからケプストラム等の音声特徴量を求めてデータベースと比較を行うという手順を踏む．また今回実験に用いた音声認識エンジンは、

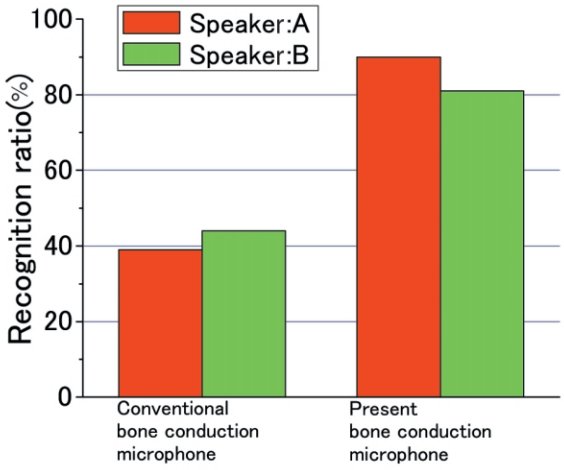


Fig. 8　Results of voice recognition test using present and conventional bone-conduction microphone

当然気導音用のものである．そこで入力された骨伝導音が音声認識されやすい音声か否かを定量的に求めるため，音声特徴量の差（=ベクトル距離）である，LPCケプストラム距離[*3]（以下CDと記す）を骨伝導音と，同時に録音した気導音の間で求めた．[2] 計算した音声区間は約0.5秒で，12kHzサンプリングの音声を，1フレーム256データポイントを128ポイントずつオーバーラップさせて，区切ってから音響分析した．

Fig. 9にあ行の各母音におけるCDを，市販骨伝導マイクロホンと今回の試作マイクロホンで、それぞれ気導音と比較して示している．比較した両音声がもし全く同一ならCDは0になり，特徴量が近い方がCDは小さい．この図を見ると，「い」と「え」においてCDが試作品より市販品の方が大きくなっていて，気導音とはそれだけ違ってきていると考えられる．それはFig. 9でも示したように，「い」「え」の第2ホルマントが他の母音より高い周波数にあり，市販の骨伝導マイクではそれらが十分検出できなかった可能性がある．一方Fig. 10でさ行の語の，CDの時間変化を示す．子音の場合，語の始まりは子音成分（この場合s音）でそのあとに母音成分が続く．図から明らかなように初めの0.15秒程度の時間が子音成分，そのあとが母音成分であり，子音成分の方でCDが大きい．また，試作マイクロホンによるCD縮小効果も子音成分の方が大きい．他の子音では，k，t，hなどの破裂音，摩擦音で母音よりもCDが大きくなる傾向が見られた．このような子音における更なる音質向上が，骨伝導音による音声認識率の向上につながると考えられる．

---

*3 LPCケプストラム距離　$CD=\frac{10}{\ln 10}\sqrt{2\sum_{i=1}^{M}(c_x(i)-c_x'(i))^2}$

ここで，$c_x(i)$ と $c_x'(i)$ はそれぞれ，基準となる音声信号と，比較する音声信号のLPC係数のケプストラム

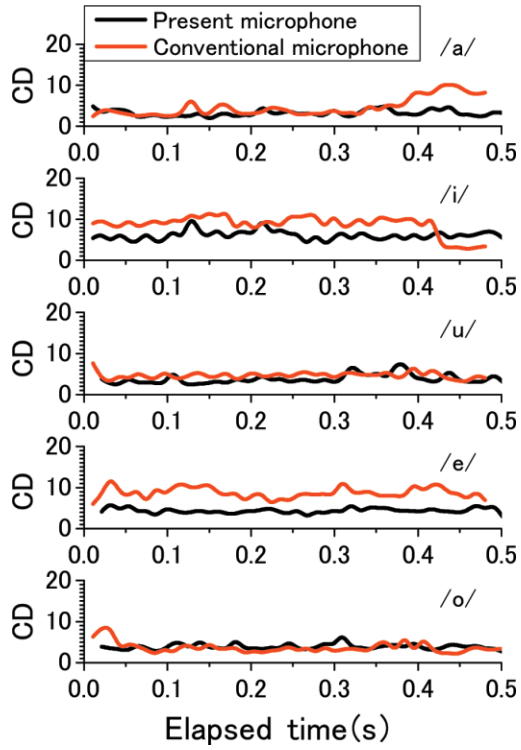


Fig. 9　Comparison of LPC cepstrum distance of Japanese vowels in present and conventional bone-conduction microphone

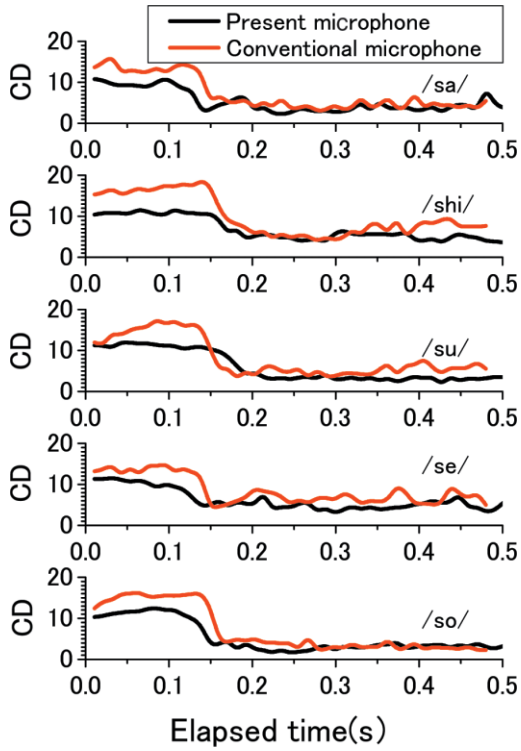


Fig. 10　Comparison of LPC cepstrum distance of Japanese consonants with /s/ sound in present and conventional bone-conduction microphone

### 4.2　今後の課題

この技術の問題点としては，まだ人による骨伝導音の音声認識率の違いが挙げられる．その原因として，大きく二つの要因を考えている．一つは，話者ごとの骨格，形状の違いなどにより，発声時に頭部を伝搬する骨伝導音が影響を受けているのではないかということである．この課題に関しては，頭部内音場シミュレーション技術を確立することで問題解決を図る予定である．もう一つの要因として，骨伝導マイクロホンの身体への装着状況の違いにより，皮膚からマイクロホンへの振動伝搬特性が異なっている可能性がある．これに関しては，身体へのマイクロホンの固定方法の検討や，小型化などにより解決していく．

## 5．おわりに

騒音環境下でも自分の声だけをS/Nよく検出できるものの，音質が悪い骨伝導音でも音声認識を可能とするために，圧電バイモルフを複数枚積層して気導音に近い骨伝導音を検出可能なマイクロホンを作製した．これを用いることで，従来40%程度だった音声認識率を80%以上にすることができた．

**補足：音響インピーダンス**

Fig. A-1に示すように媒質Ⅰから媒質Ⅱへ音波が境界面に垂直に入射した場合，媒質Ⅰの密度を$\rho_1$，音の速度を$c_1$，媒質Ⅱの密度を$\rho_2$，音の速度を$c_2$とし，入射する音の強さをIi，反射する音の強さをIrとすると透過率τは

$$\tau = \frac{I_i - I_r}{I_i} = \frac{4\rho_1 c_1 \rho_2 c_2}{(\rho_1 c_1 \rho_2 c_2)^2}$$

で表すことができる．二つの媒質の音響インピーダンス$\rho c$が非常に異なる場合，すなわち$\rho_1 c_1 \ll \rho_2 c_2$の場合には，上式よりτ　0となる．一般的に，気体の音響インピーダンスは～$10^2$（Ns/m$^3$），固体の音響インピーダンスは～$10^7$（Ns/m$^3$）である．以上のことから，空気中（気体）を伝わってきた騒音（音）が，人体（固体）に透過する率は極めて小さい．

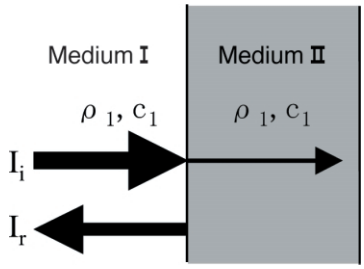


Fig. A-1 Sound propagation from gaseous medium to solid medium

## ＜参考文献＞

1) RAY D. KENT, CHARLES READ,（監訳）荒井　隆行，菅原　勉：音声の音響分析，海文堂（2000）
2) 島村　徹也：MATLABプログラム事例解説Ⅰ音声通信～特徴抽出と雑音低減～，トリケップス（2001）, pp. 31-59.

## ＜著　者＞

山田　芳靖
（やまだ　よしやす）
基礎研究所
工学博士
音声HMIの研究に従事

土方　啓暢
（ひじかた　よしまさ）
基礎研究所
音声HMIの研究に従事