

低域通過フィルタを用いた気導音からの自己聴取音の再現

落合 裕貴¹ 丹治 寛樹¹ 村上 隆啓² 松本 直樹²

概要: ある発話者の話す声をマイクロホンで録音するとき、発話者自身が聴取する音声と録音される音声は異なる。そのため、発話者が録音された音声を聴取すると違和感を覚える。この違和感は、録音された音声に骨導音が含まれていないことが原因とされている。そこで、この違和感を軽減するために、マイクロホンで観測した気導音から骨導音を推定し、観測信号と足し合わせることで、自己聴取音を再現する。本稿では、気導音から骨導音を推定する手段として、気導音に低域通過フィルタを適用する方法を提案する。また、提案法によって再現した自己聴取音と、発話者の気導音を比較する主観評価実験によって、提案法が違和感の軽減に対して有効であることを確認する。

Reproduction of autophonic production from air-conducted sound using low-pass filters

YUKI OCHIAI¹ HIROKI TANJI¹ TAKAHIRO MURAKAMI² NAOKI MATSUMOTO²

1. はじめに

人間は音声コミュニケーションを行う際、対話相手の音声を聴取するだけでなく、発話者自身の音声(自己聴取音)も聴取している。発話者は、目標としている音声の音響特徴や調音特徴に近づけるため、自己聴取音を参考に調音運動の制御を行う。これを聴覚フィードバック [1] と呼び、人間が正常に発声するために必要な機能である。聴覚フィードバックを用いた研究として、吃音障害を軽減する研究 [2] や、発話と音声聴取の間に生じる遅延に関する研究 [3] 等が行われている。このような研究を行うにあたり、マイクロホンを用いて発話者の気導音を観測し、その音声を研究目的に合わせて処理する方法がよく用いられる。しかし、この方法は発話者の自己聴取音とマイクロホンで観測された音声の間に差異が生じることが確認されている [4]。この差異は、聴覚フィードバックに関する研究に限らず、たとえば、ビデオカメラを用いて自分自身が発話している動画を視聴した場合にも生じるため、日常生活を送るうえで体験することもあり、一般的に広く知られているものである。

図 1 に自己聴取音認知のメカニズムを示す。声帯の振動

が発話者の口腔から放射され空気振動となり、その空気振動が発話者の外耳道内に入る。この音声の知覚経路を気導音と呼ぶ。一方、声帯の振動が体内組織を通過し、頭蓋骨を振動させることで音声を知覚する経路や、体内組織を通過した振動が外耳道の空気を振動させることで音声を知覚する経路が存在する [5]。この経路を骨導音と呼ぶ。骨導音と呼ばれる経路の中でも、体内組織を通過した振動が外耳道の空気を振動させることで音声を知覚する経路が聴覚に最も関与が大きいことがわかっており [5]、本稿ではこの経路を骨導音として用いる。発話者は骨導音と気導音を同時に聴取することで自己聴取音を認知している。しかし、マイクロホンで観測された発話者の音声は気導音のみを含み、発話者の骨導音を含まない。これがマイクロホンで観測された発話者の音声と自己聴取音の間に生じる違和感の原因と考えられている [6]。

今までにも気導音に骨導音の成分を合成させることで自己聴取音を再現しようとする研究は行われており [6], [7]、これらの研究はどちらも加速度計を被験者に装着することで骨導音を観測し、それをマイクロホンで観測した気導音に合成させている。この測定方法は、発話者にとって負担となることや、観測を行うために必要な装置が増えることなどの問題

¹ 明治大学大学院理工学研究科電気工学専攻

² 明治大学理工学部電気電子生命学科

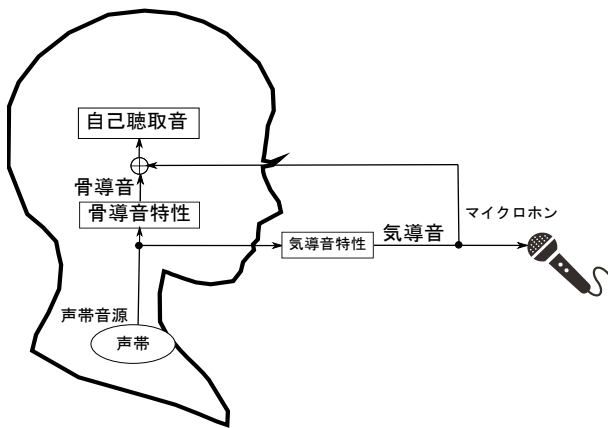


図 1 自己聴取音認知のメカニズム

点が生じる。そこで本研究では、マイクロホンを用いて発話者の気導音のみを観測し、そこから骨導音を推定することで、発話者の自己聴取音を再現し、自己認知音と気導音の間に生じる違和感の軽減を目指す。マイクロホンを用いて観測された気導音のピッチを変えさせたり、周波数成分を変化させたりすることで自己聴取音を生成する研究は過去に報告されている [4], [8] が、骨導音を気導音から推定する方法は報告されていない。

本稿では、まず、マイクロホンと加速度センサを用いて発話者の気導音と骨導音の間に生じる関係を調査する。そこで確認された気導音と骨導音の関係をを用い、気導音から骨導音を推定するための骨導音再現フィルタを決定する。このとき、マイクロホンで観測された気導音と加速度センサで観測された骨導音では、マイクロホンの入力と加速度センサの入力利得が異なるため、気導音と骨導音の間の振幅の関係を決定することができない。そこで、自己聴取音に含まれる気導音と骨導音の割合についての主観評価実験を行い、自己聴取音に対する骨導音の含有割合について調査をする。

2. 自己聴取音を再現するシステムの提案

2.1 気導音と骨導音の関係

気導音を用いて骨導音を推定するために、気導音と骨導音の関係を明らかにする必要がある。そこで、気導音と骨導音を同時に観測し、その間に見られる関係について考察する。図 2 のようにマイクロホンと加速度センサを接続し、気導音と骨導音を同時に観測する。使用する加速度センサおよびマイクロホンを表 1 に示す。本稿では、サンプリング周波数 48000Hz とする。発話者は、日本語の 5 母音 (/a/i/u/e/o/) を 1 秒間発声する。観測した気導音および骨導音の時間波形の例を図 3 に示す。気導音と骨導音の間には観測される波形に時間差 (遅延) が生じていることがわかる。

この遅延は、気導音と骨導音の知覚経路の差によって、空気中の音速と固体や液体中の音速が異なることから生じ

表 1 気導音・骨導音観測の使用機器

使用機器	製造会社	型番
マイクロホン	DPA MICROPHONES	SC4060
加速度センサ	小野測器	NP-3211
センサアンプ	小野測器	SR-2210
オーディオインターフェース	Roland	OCTA-CAPTURE

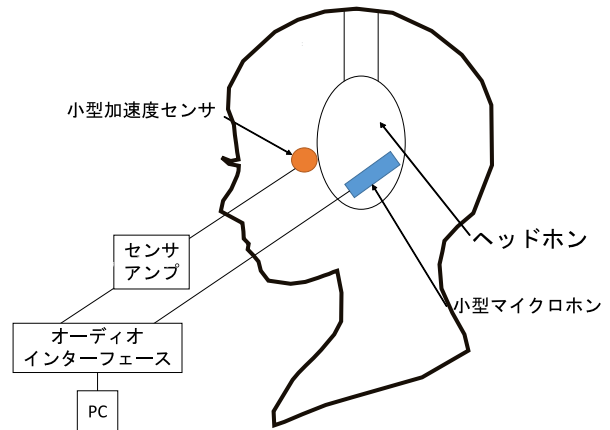


図 2 骨導音と気導音の同時観測方法

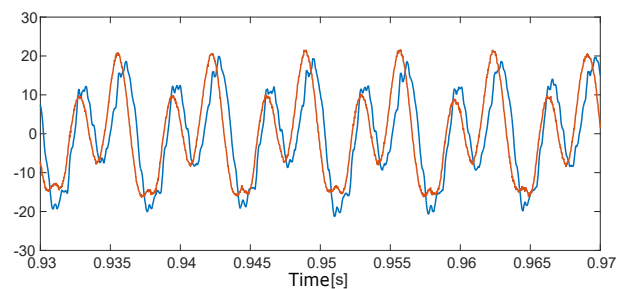


図 3 気導音と骨導音の時間波形 (母音/u/)
青: 気導音 赤: 骨導音

ると考えられる。遅延の大きさは被験者、母音により異なり、遅延が小さい被験者で 0.3ms、大きい被験者で 0.8ms 程度であった。人間が知覚できる音声の遅延は 3ms 以上であるという調査が行われており [3]、発話者の骨導音と気導音の間に生じる遅延は認知できない程の遅延時間の範囲に含まれるため、発話者自身が自身の気導音と骨導音を聞き分けていることは考えられない。一方、自己聴取音を気導音と骨導音の合成と考えた場合、自己聴取音に対して遅延によるノッチが発生していることが考えられる。

次に、図 4 に気導音と骨導音のパワースペクトルの例を示す。骨導音は気導音と比較し、低域の周波数成分が強調され、高域の周波数成分が減衰している。また、図 5 に 500Hz から 2000Hz におけるパワースペクトルを示す。骨導音のパワースペクトルは 800Hz 付近から緩やかに減衰をはじめ、気導音のパワースペクトルの 1800Hz 付近に見られるようなピークが消失している。そこで、緩やかに減衰する周波数特性を持つ低域通過フィルタを用いることで、気導音から骨導音を推定する。

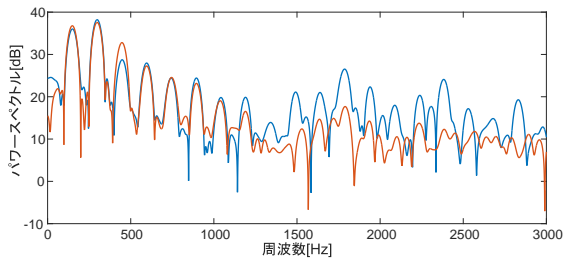


図 4 気導音と骨導音のワースペクトル (母音/u/)
青: 気導音 赤: 骨導音

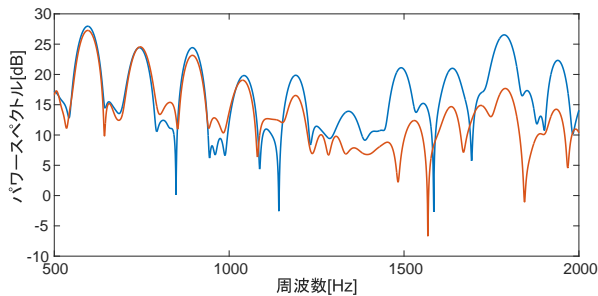


図 5 気導音と骨導音のワースペクトル (母音/u/, 500Hz-2000Hz)
青: 気導音 赤: 骨導音

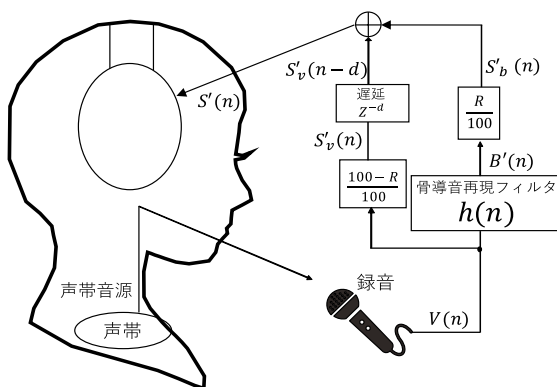


図 6 自己聴取音を再現するシステム

2.2 自己聴取音を再現するシステム

気導音から自己聴取音を再現する方法を検討する。本稿では、自己聴取音を再現するために、気導音から骨導音を推定し、推定された骨導音と気導音を合成する。マイクロホンで観測された気導音から骨導音を推定するとき、骨導音の振幅の決定ができない。そこで、自己聴取音を再現するために、推定された骨導音と気導音を合成する割合を考慮する必要がある。自己聴取音を再現に用いるシステムを図 6 に示す。このシステムでは、マイクロホンで観測された気導音に低域通過フィルタを用い、疑似的な骨導音を推定する。疑似的な骨導音に気導音を、適切な割合と遅延を与えて足し合わせることで、自己聴取音を再現する。図中の変数について表 2 で説明する。変数 R , d , フィルタの特性 $h(n)$ を決定することで、自己聴取音を再現する最適なシステムを求める。

表 2 自己聴取音を再現するシステムの変数

$V(n)$	観測された気導音
$h(n)$	3.1 節で得られるフィルタの特性
$B'(n)$	3.1 節で推定される骨導音
R	自己聴取音の中に含まれる骨導音の割合 [%]
$S'_b(n)$	再現した自己聴取音の中に含まれると推定される骨導音
$S'_v(n-d)$	再現した自己聴取音の中に含まれると推定される気導音
d	気導音と骨導音の間に生じる遅延
$S'(n)$	再現された自己聴取音

表 3 フィルタのパラメータ

	カットオフ周波数 [Hz]	フィルタ次数
①	600	4 次
②	800	4 次
③	1000	4 次
④	1200	4 次
⑤	600	2 次
⑥	800	2 次
⑦	1000	2 次
⑧	1200	2 次

3. 自己聴取音の再現に関する主観評価実験

自己聴取音の再現をするシステムを決定するために必要な変数を求めるため、主観評価実験を行う。

3.1 低域通過フィルタの特性を決定する主観評価実験方法

骨導音の推定に用いるフィルタの特性について検討する。本研究では、2.1 節で得られた結果をもとに、周波数特性が緩やかかつ、1800Hz 以上の周波数成分を大きく減衰させる低域通過フィルタを用いる。本稿では、周波数特性が緩やかなフィルタとして、フィルタ次数の低いバターワースフィルタを用いる。表 3 に示すパラメータの低域通過フィルタを発話者の気導音に用い、最も発話者の骨導音に近い音声を得られる条件を主観評価実験によって決定する。主観評価実験の手順を以下に示す。

- (1) ヘッドホンと加速度センサを図 2 のように発話者に装着する。その際、ヘッドホンの左耳部分に小型マイクロホンを装着する。
- (2) 発話者は、測定者からの合図を基準に日本語母音/a/を 2 秒間発話し、その気導音と骨導音を観測する。
- (3) 測定者は、観測された発話者の気導音に表 3 の①の条件のフィルタを適応し、音声を生成する。
- (4) 測定者は、(3) で生成した音声と発話者の骨導音を、0.5 秒の間隔を空けて発話者のヘッドホンに呈示する。発話者は、ヘッドホンから出力された 2 つの音声を聴取し、双方の間に差異があるかを確認し、「差異がわかる」「差異がわからない」の 2 択で回答する。
- (5) 表 3 の②～⑧の条件について同様に (3), (4) を行う。
- (6) 日本語母音/i/u/e/o/について同様に (2)～(5) を行う。

3.2 低域通過フィルタの特性を決定する主観評価実験結果

フィルタ次数が4次の音声に関してどちらも骨導音に近い音声が得られているという評価を得た。一方、フィルタ次数が2次の音声に対しては、骨導音に近い音になっていないという評価を得た。カットオフ周波数を変化させた場合、600Hzの時は音声がこもりすぎて骨導音に近くないという評価を、1200Hzの時は音声が明瞭すぎて骨導音に近くないという評価を得た。カットオフ周波数が800Hzと1000Hzについて主観評価実験の結果に差異は無く、どちらも骨導音に近いという評価を得た。2.1節では、骨導音のパワースペクトルで気導音のパワースペクトルの1800Hz付近に見られるようなピークが消失しているという結果が得られており、気導音と骨導音の音色の差は骨導音における1500Hz以上の周波数成分の減衰により起きると考えられる。そこで、主観評価実験の結果に差異が生じないカットオフ周波数が800Hzと1000Hzについて、1500Hz以上の成分をより減衰させるため、カットオフ周波数800Hzの4次のバターワースフィルタを用いる。

3.3 自己聴取音の再現に対する主観評価実験方法

2.2節で提案したシステムを用い、気導音から自己聴取音の再現を行い、再現された自己聴取音について的主観評価実験を行った。主観評価実験の手順を以下に示す。

- (1) 被験者にヘッドホンを装着する。その際、ヘッドホンの左耳部分に小型マイクロホン装着する。
- (2) 被験者は、測定者からの合図を基準に、日本語母音のみで作られた短文「いい青い家/iiaoiie/」を発話する。
- (3) 測定者は、発話された音声を2.2節のシステムに従い、骨導音の含有割合 R を変化させて被験者のヘッドホンに呈示する。音声の呈示は発話し終えてから0.5秒後とし、被験者の発話と重ならないようにした。被験者は、ヘッドホンから出力された音声を聴取し、自己聴取音に近いかという質問に対し「近い」「近くない」「どちらともいえない」の3択で回答する。
- (4) 自己聴取音の中に含まれる骨導音の割合を0%~100%まで5%ずつ変化させ、(2)(3)を繰り返す。
- (5) 自己聴取音に近いという評価がされた骨導音の含有割合の音声に対し、0ms~約1.04ms(0point~50point)まで約0.104ms(5point)ずつ遅延を加え、(2)(3)を繰り返す。自己聴取音に近いという評価がされた骨導音の含有割合が4種類以上存在した場合、その中で最も骨導音の含有割合が小さいもの、最も大きいもの、中間値をとるものの3種類について実験を行う。

被験者は20代の男性9名で、主観評価実験を行う日の体調や気分により結果が異なることが予想されるため、別日に同様の実験を行う。

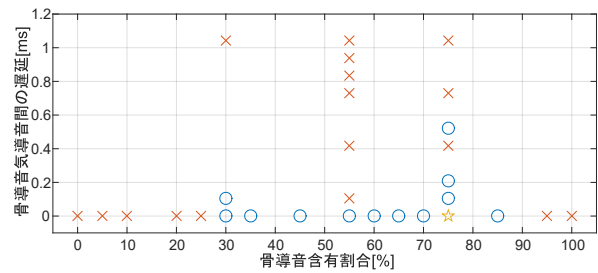


図7 被験者4の自己聴取音の再現に対する主観評価実験結果
 (○:自己聴取音に近い ×:自己聴取音に近くない
 ☆:最も自己聴取音に近い)

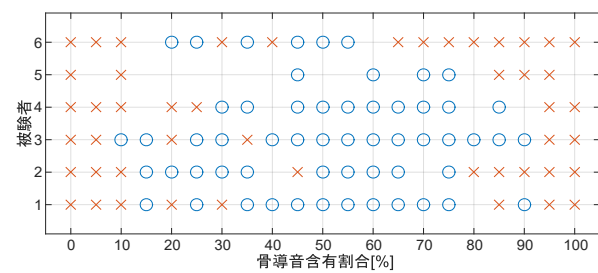


図8 被験者全体の自己聴取音の再現に対する主観評価実験結果
 (○:自己聴取音に近い ×:自己聴取音に近くない)

3.4 自己聴取音の再現に対する主観評価実験結果

ある被験者の実験結果を図7に示す。遅延を考慮しない実験環境下において、骨導音の含有割合が30%~75%および85%の場合に自分の声らしいという評価が得られた。一方、骨導音の含有率が0%~25%の音声について、音声が明瞭すぎる、自分の声ではないような違和感を覚える、という意見が得られた。また、骨導音の含有率が90%~100%の音声について、機械で変換したのがわかる、くぐもった音声で自分の声とは程遠い、という意見が得られた。図8はすべての被験者を対象にした、骨導音の含有割合に対する評価である。骨導音の含有割合が95%を超えたとき、すべての被験者から「出力された音声が自分の声よりもこもっている、ピッチが低く聞こえる」という意見が多く、評価が低くなった。また、骨導音の含有割合が10%以下の場合も、「出力された音声が明瞭すぎる、ピッチが高く聞こえる」という意見を得た。しかし、それ以外の音声に関して、被験者により評価に差異が生まれた。また、骨導音の含有割合が50%~60%のとき、自己聴取音と比較して違和感があると回答した被験者はいなかった。また、遅延を考慮した実験では、0.7ms以上の遅延が生じると評価は悪化し、遅延がない場合と比べてよい評価を得た音声はなかった。しかし、0.7ms未満の遅延に対する評価は被験者ごとに異なり、遅延のある音声に対しても遅延のない音声と同様に自己聴取音に近い評価をする人と、自己聴取音と離れている評価をする人がいた。今後さらなる調査を行い、遅延を考慮するべきか検討する必要がある。

4. まとめ及び今後の課題

本稿では、発話者から観測された気導音のみを用いて、発話者の骨導音を推定し、推定された骨導音と観測信号である気導音を足し合わせることで、自己聴取音の再現を行った。本稿で提案したシステムにより、発話者自身の気導音に対する違和感を軽減できることを確認し、その時の骨導音の含有割合の範囲を得ることができた。一方、推定された骨導音と観測された気導音の間に生じる遅延に関する主観評価実験の結果は被験者ごとに大きく異なっており、今後も継続的な調査が必要になると考えられる。本稿では、発話者の対象を20代男性に限定しており、それ以外の年齢層や性別の発話者について考慮していない。男性の骨導音は女性の骨導音に比べ明瞭度が高いという報告がされており [9]、発話者の性別や年齢層によって、骨導音を推定するためのフィルタの特性は適宜変更する必要があると考えられる。また、自己聴取音の再現を行う際、本稿では日本語母音のみを用いた文章で実験を行った。しかし、日常生活を行うにあたり、日本語母音のみの文章を用いて生活することは考えられない。子音には、破裂音、摩擦音、鼻音等の種類が存在し、子音の種類によって骨導音の振幅の増減が発生するという報告がされている [10] ため、子音を含めた文章について今回提案したシステムが有効であるか調査を行い、システムやパラメータの設定を再考する必要がある。

参考文献

- [1] Perkell, J. S.: Movement goals and feedback and feedforward control mechanisms in speech production, *Journal of Neurolinguistics*, Vol. 25, No. 5, pp. 382–407 (2012).
- [2] 岡崎俊太郎, 武田湖太郎: 吃音と聴覚フィードバック, *脳科学とりハビリテーション*, Vol. 14, pp. 35–40 (2014).
- [3] 碓田猛真, 中村陽祐, 福元儀智, 長谷川賢作, 北野博也: ディレイタイムの認知閾値, *AUDIOLOGY JAPAN*, Vol. 46, No. 5, pp. 465–466 (2003).
- [4] Nakayama, I.: Voice timbre in autophonic production compared with that in extraphonic production, *Journal of the Acoustical Society of Japan (E)*, Vol. 18, No. 2, pp. 67–71 (1997).
- [5] 細井裕司: 軟骨伝導聴覚の基礎と補聴器等への応用, *日本耳鼻咽喉科学会会報*, Vol. 117, No. 2, pp. 146–147 (2014).
- [6] 中井孝芳, 高尾諭司: 発声者自身の音声の知覚経路, *電子情報通信学会技術研究報告*, Vol. 101, No. 73, pp. 15–22 (2001).
- [7] 森 幹男, 吉田千夏, 荻原慎洋, 谷口秀次, 高橋謙三: 自己聴取音に占める気導音と骨導音の割合の推定, *電気学会論文誌 C*, Vol. 127, No. 8, pp. 1268–1269 (2007).
- [8] 伏見遠平, 吉田成朗, 鳴海拓志, 谷川智洋, 廣瀬通孝: 自分の声を知り, コントロールするための「自分声フィルタ」の提案, *エンタテインメントコンピューティングシンポジウム 2016 論文集*, Vol. 2016, pp. 100–105 (2016).
- [9] Hodges, M. L. and McBride, M. E.: Gender differences in bone conduction auditory signal processing: Communication equipment design implications, *International*

Journal of Industrial Ergonomics, Vol. 42, No. 1, pp. 49–55 (2012).

- [10] 加藤達也, 島村徹也: 骨導音声における子音-母音間の相対振幅, *電子情報通信学会技術研究報告*, Vol. 112, No. 369, pp. 71–74 (2012).