対話型遺伝的アルゴリズムを用いた
自己聴取音合成のための効率的なパラメータ推定*

田中尚輝，増田尚建，齋藤大輔，峯松信明 (東大・工学系)

1　はじめに

録音された自分の声を聞いてみると，自分で話しながら聞く自分の声 (以下，自己聴取音と呼ぶ) と違う印象を受けることがある [1, 2]．それは，録音された音声が空気を振動させ鼓膜を通じて伝達される気導音だけなのに対して，自己聴取音は気導音に加え，発声器官から体内の器官を振動させ内耳に伝達される体内音が混入するためである．体内音は一般に「骨導音」と呼ばれるが，[3, 4] によれば，自己聴取音の生成に様々な経路が関与していることが示唆されており，[5] ではどこの骨（部位）の振動なのかを明示させ，耳骨導音，喉骨導音を，専用の振動検出器（耳骨導マイク，喉骨導マイク）で計測している．[3, 4] で示された各種経路の振動現象を個別に計測することは不可能であるが，体内を通して内耳に至る振動の総和（＝自己聴取音から気導音を差し引いたもの），を発声者自身が聴取することは可能である．防音用イヤーマフを装着して気導音を遮断して話せばよい．[5] ではこれを（個別の骨導音と区別して）体内音として定義し，気導音，耳骨導音，喉骨導音の 3 種類の音源を使って，体内音及び自己聴取音の合成を検討している．個々の音源にフィルターをかけ，重み付きで足し合わせて合成しているが，フィルター特性及び重みを，個別に，手動で決定しており，時間がかかる上に，得られたフィルター特性，重みの最適性にも問題がある．本研究では，対話型遺伝的アルゴリズム（Interactive Genetic Algorithm, IGA）を用いてこれら問題の解決を試みたので報告する．

2　先行研究

2.1　耳・喉骨導音を用いた自己聴取音の合成 [5]

気導音，耳骨導音，喉骨導音を同時収録し，**図 1** に示す方式で自己聴取音を合成している．三種類の音源に対して，個別のフィルタを適用し，それらの重みづけ和として体内音を合成する．その後，体内音と気導音を再度重みづけで足し合わせて自己聴取音を合成している．三種類のフィルタ（$H_a(\omega)$，$H_e(\omega)$，$H_t(\omega)$）は，各々のフィルタ出力が，個別に，イヤーマフを装着して聴取できる体内音に近づくよう，手動で設計している．フィルタ設計は，**図 2** に示す 20 チャネル

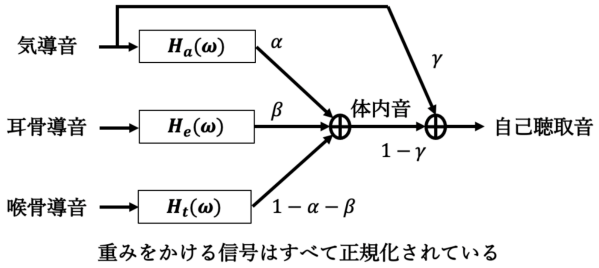

Fig. 1　自己聴取音推定を段階的に求める手法 [5]

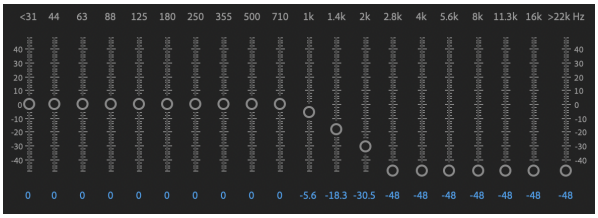

Fig. 2　[5] で使われていたイコライザー

のグラフィカルイコライザーを使用している．基本的に LPF が構成されるが，利用者に各チャネル強度の増幅・抑制によって音がどう変わるのかを学ばせた上で，設計させていた．$H_a(\omega)$，$H_e(\omega)$，$H_t(\omega)$ を個別に設計後，イヤーマフを用いた聴取実験によって，重み α，β を決め，その後，改めて，体内音と気導音の重み γ を聴取実験によって決定してる．

[5] によれば良好な体内音，自己聴取音が得られていたが，下記の問題が指摘されている．1) フィルタや重みの設計が，個別に行われおり，全体最適化となっていない．2) グラフィックイコライザの操作は，信号処理に関する専門知識のない利用者（実際，中学生を相手にデータ収集している）の場合，各チャネルの意味が分からないまま操作している可能性がある．3) フィルタ，重み設定を全て聴取実験を通して行っているが，その作業に時間がかかる．本研究は，これらの問題を下記に示す IGA を用いて解決を図る．

2.2　対話型遺伝的アルゴリズム [6]

遺伝的アルゴリズムとは，解の候補を遺伝子で表現した個体を複数用意し，個体の適応度を求め，適応度の高い個体から優先的に選択して交叉・突然変異などの操作を行い新しい個体を用意する，という一連の流れを繰り返しながら最適解を探索する手法である．一般的な遺伝的アルゴリズムでは最適解にどれだけ近いかを示す定量的な指標が存在することが多く，

*Efficient parameter estimation for synthesizing one's own voice with Interactive Genetic Algorithm, by Naoki Tanaka, Naotake Masuda, Daisuke Saito, Nobuaki Minematsu (UTokyo)

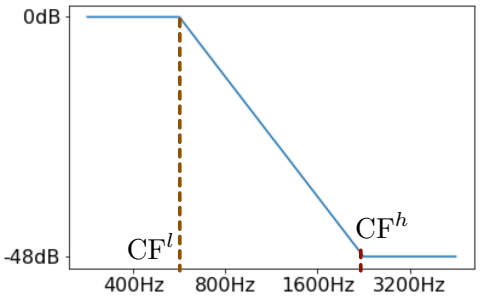

Fig. 3　体内音化 LPF における CF^l と CF^h

その場合は適応度の評価や交叉・突然変異など新しい個体を生成する作業を含めて全て自動で処理できる．

しかし，自己聴取音合成用のパラメーター推定においては，[5, 7–10] から，最適値が話者依存になることが示されている．また，真の自己聴取音や体内音は wav ファイルとして外部保存できず，合成音声（模擬音声）の評価は，話者本人の発声や聴取によって可能という特殊性もある．そのため，話者本人に聴取実験による評価を行わせ，その評価値を元に新しい個体を遺伝的アルゴリズムにより決定する対話型遺伝的アルゴリズム（IGA）を採用した．IGA を用いて主観評価を最適化する先行研究として [11, 12] がある．

[12] では補聴器のパラメータ設定を IGA を用いて，装着者毎に，効率的に実施している．聴覚障害の程度は人によって異なるため [13]，患者ごとに補聴器を調整する必要がある．本研究でも [12] に倣い，各種パラメータの効率的な最適化を図る．

3　提案手法

図 1 に示す体内音・自己聴取音合成方法において，効率的なパラメータ推定を IGA で検討する．[5] で求まった各話者のフィルタ特性，重みのレンジ，更に，筆者らの予備的検討を元に，IGA におけるパラメーターは以下の 9 つとした．

1) 低域通過フィルタ特性パラメータ（**図 3** 参照）
- $H_a(\omega)$ の cut-off 周波数（CF$_a^l$ と CF$_a^h$）
- $H_e(\omega)$ の cut-off 周波数（CF$_e^l$ と CF$_e^h$）
- $H_t(\omega)$ の cut-off 周波数（CF$_t^l$ と CF$_t^h$）

2) 波形重畳用の重み（**図 1** 参照）
- 気導音，耳・喉骨導音を体内音化する重み（α，β）
- 気導音と体内音を自己聴取音化する重み（γ）

提案手法のシステム（以下 IGA システムと呼ぶ）の概念図を**図 4** に示す．まず，ランダムな初期解 8 個に基づく合成音を被験者に提示し，どれだけ自己聴取音/体内音に近いかを 1〜5 点で主観評価させる．その評価を元に，8 個の中で最良の合成音 1 つを次世代に残す（エリート選択）．最良合成音が複数存在した場合は，その中からランダムに 1 つ選出する．

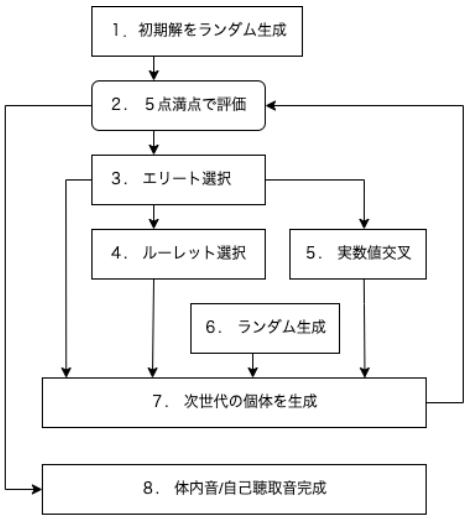

Fig. 4　提案する IGA システム

Fig. 5　自己聴取音聴取・評価用 GUI

次に，未選択の 7 個の音 $i_1 \ldots i_7$ の評価値を $f_1 \ldots f_7$ とし，$p_i = f_i / \sum_{k=1}^{7} f_k$ で定義される確率 p_i を用い，1 つをルーレット選択により選んで次世代に残す．即ち p_i を選択優先度としたランダム選択となる．

上記で選ばれた 2 個体のパラメーターを元に，探索領域を 2 つのパラメーターの近傍に絞った上でランダムに 2 つ合成音を生成する（実数値交叉，BLX-α 法）．BLX-α 法における探索領域の広さを決定する係数である α の値は 0.5 とした．

最後に，探索領域の制限をより緩めて，再度ランダムに 4 つ生成する．これら 4 ステップで次世代の個体（8 個）を決定し，自己聴取音/体内音を合成して再度被験者に主観評価させる．以上を繰り返し，被験者が満足できる自己聴取音/体内音を見つけるまで繰り返す．満足した合成音のスコアは，MOS 相当の評価値となる．なお，実験の遂行を円滑にするため，**図 5** に示す GUI を作成して聴取・評価実験を行った．


Fig. 6　耳骨導マイクと喉骨導マイク

Fig. 7　防音用イヤーマフと小型スピーカ

4　実験

4.1　使用した機器と音声提示環境

図 1 に示す 3 種類の音源は，一般的なコンデンサーマイク，イヤホン型の耳骨導マイク (TEMCO EM20N-T3)，咽喉型の喉骨導マイク (TEMCO TM80N-T) を用いて収録した（**図 6** 参照）．防音用イヤーマフ (3M PELTOR X5A) を用い，被験者自らが発声することで自身の体内音を確認できるようにした．このイヤーマフは内面に吸音材としてウレタンが貼られており，イヤーカップ内に生じる定常波を低減させており，体内音を確認するための適したデバイスとなっている（**図 7** 参照）．更に，小型無線スピーカー (Bearoam Bluetooth 5.1 耳掛け式ワイヤレスイヤホン) をイヤーマフ内に装着し，模擬体内音をイヤーマフ装着時にも提示できるようにした．こうすることで模擬体内音聴取時に，自らが発声することで，真の体内音を確認できる．

4.2　被験者と読み上げ原稿

成人男性 6 名，女性 2 名が参加した．体内音化，自己聴取音化のための音源として，全ての清音が一つずつ含まれる現代いろは唄の「クレヨン」を用いた．気導音マイクと異なり，骨導マイクは皮膚に接着させるため，収録中にマイクの位置がずれると収録音声の特性が変わることがある．必要に応じて接着用のテープなどを使って固定した．収録後，Adobe Audition 2021 を用いてノイズの削除，音量の正規化を行った．

4.3　実験手順

以下の 3 手法で，被験者毎に自己聴取音を作成し，その後，3 種類の模擬自己聴取音を真の自己聴取音に近い順に順位づけさせた．IGA システムを使う場合の具体的な手順は**第 3 節**に示した通りである．更に，タスクの簡便さ，所要時間についても評価した．

Table 1　気導音・体内音間の重み γ

気導音 (γ)	0	0.2	0.4	0.6	0.8	1
体内音 ($1-\gamma$)	1	0.8	0.6	0.4	0.2	0

Table 2　主観評価実験の結果
(○: 一番似ている，△: 次に似ている，×: 一番似ていない)

被験者	手法 1	手法 2	手法 3	MOS
M01	×	△	○	4.7
M02	×	△	○	5.0
M03	×	○	△	5.0
M04	△	○	×	4.0
M05	△	○	×	5.0
M06	×	△	○	4.6
F01	×	△	○	5.0
F02	△	○	×	5.0

手法 1　先行研究 [5]，**図 2**，**図 1** と同様，20 チャネルのイコライザーを用いてフィルターと重みを段階的に求めて，自己聴取音を推定する手法

手法 2　IGA システムで自己聴取音を推定する手法

手法 3　IGA システムで体内音を推定した後，気導音と模擬体内音を 6 種類の重みで合成し，一番良い音を自己聴取音とする 2 段階推定法

手法 2 と 3 の違いであるが，2 は**第 3 節**で示した 9 つのパラメータ全体の準最適化を狙ったものであり，手法 3 は，**図 1** における体内音の合成と重み γ の推定の独立性を仮定し，前者を一旦求めてから，改めて後者を求める 2 段階推定となっている．手法 3 の気導音と体内音の重み γ は，**表 1** を用いた．

4.4　実験結果と考察

評価者毎に 3 手法で模擬自己聴取音を作成し，各々を比較して，順位づけさせた．結果が**表 2** である．MOS は，最終的に選択された合成音声（○）の評価値である．全員が手法 2 は手法 1 に比べて自己聴取音に近いと答えていることから，提案手法が，模擬自己聴取音の品質を向上させていることが分かる．[5] では，フィルタ設計，重み推定など，処理過程の多くにおいて独立性を仮定しており，IGA の全体（準）最適化の効果は非常に大きいと考えられる．

手法 2 と手法 3 であるが,「最も似ている」と答えた被験者数は同数であるが,「次に似ている」と答えた被験者数は手法 2 の方が多い．結果的に，全体の最適化を図った手法 2 の方が優れていると判断されるが，被験者 4 名が手法 2 より手法 3 の方が優れいると判断している．体内音推定までの過程と，気導音と体内音を合成する重み推定の過程の独立性は比較的高いと解釈できる．[3, 4] は，様々な部位から内耳に振動が伝わる様子を説明しているが，各々の経路のう

ち，どの経路が支配的なのか，どの経路とどの経路は独立性が高いのか，については話者に依存すると推察される．なお，IGA に起因することであるが，パラメーター推定に際してランダム制御を用いており，試行を重ねる度に（わずかに）異なる自己聴取音が生成される．このばらつきを抑えるためのアルゴリズムの提案は今後の検討事項である．

タスクの簡便さ（所要時間）について聞いたところ，8 名全員が手法 1 より，手法 2，3 の方が明らかに負担が軽く，所要時間も短いと回答した．手法 1 では 1 時間以上要したのに対し，手法 2, 3 ではどの被験者も 3〜5 世代で終了し，所要時間は最大 30 分であった．手法 2, 3 では IGA システムの部分に大差はなく，後者は，IGA システム利用の後に，合成音声を 6 個聞く必要があり，若干余分に時間を要していた．

特に 20 チャネルのイコライザーを操作するのは，たとえ信号処理の知識があったとしても，音響機器の扱いに関する知識がないと，作業の信頼性（再現性など）も低くなると思われる．今回の被験者は全員成人であるが，別途，中学一年生を対象とした実験も行っており，特に「IGA システムの利用は専門知識が不要である」特性は大きく貢献できるだろう．

なお，イヤーマフの利用であるが，模擬体内音を真の体内音とを比較する場合は非常に適した提示装置であった．この提示条件をそのまま，模擬自己聴取音の評価の時にも使ったが，真の自己聴取音を確認する場合，イヤーマフは当然邪魔になる（外す必要がある）．この作業が負担となっていたことは否めず，より簡便な実験環境を求める場合，自己聴取音の評価においては，外部スピーカの利用を考える必要がある．

5　おわりに

本研究では，対話型遺伝的アルゴリズムを用いて心的負担を減らしながら高精度に自己聴取音を推定する方法を提案した．IGA を用いると複数のパラメータの（準）全体最適化が可能となる．体内音と気導音の独立性の有無を考えて 2 通りの手法を検討したが，いずれも先行研究の手法を上回る模擬自己聴取音を，より短時間の試行で合成することができた．この研究により，自己聴取音合成技術の活用方法についての研究がより盛んになると期待される．

自己聴取音合成技術の活用として，例えば外国語学習が考えられる [14]．母語話者のモデル音声の話者性を，学習者の話者性に変換する手法が検討されている（Cross-lingual voice conversion[15]）．この手法を用いると，学習者自身の声色のモデル音声が取得できるが，それは気導音・録音声であって，学習者が聞く自分の声ではない．この合成音声を自己聴取音化すれば，一番真似しやすいモデル音声となる可能性がある．他にも，ゲームや映像作品への応用が考えられる．[16] ではゲーム内でのアバターの声が自身の声に近いほどゲーム内のタスクの完成度，没入感が向上したと報告されている．この実験では気導音が用いられていることから，自己聴取音を用いた場合ではよりよい結果が得られると考えられ，作品中の表現の一つとしての自己聴取音の活用も期待される．

参考文献

[1] C. L. Rousey ¡5¡5*et al.*, “Recognition of one’s own voice,” *Journal of Personality and Social Psychology*, 6, 4, 464–466, 1967.

[2] A. J. Weston *et al.*, “Voice confrontation in individuals with normal and defective speech patterns,” *Perceptual and Motor Skills*, 30, 1, 187–190, 1970.

[3] S. Stenfelt *et al.*, “Bone-conducted sound: physiological and clinical aspects,” *Otology & Neurotology*, 26, 6, 1245–1261, 2005.

[4] S. Stenfelt, “Acoustic and physiologic aspects of bone conduction hearing,” *Implantable Bone Conduction Hearing Aids*, 71, 10–21, 2011.

[5] R. Chen *et al.*, “Acoustic simulation of body-conducted speech and its use to convert one’s recorded voices to one’s own voices,” *Proc. APSIPA*, 821–828, 2021.

[6] H. Takagi, “Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation,” *Proc. IEEE*, 80, 9, 1275–1296, 2001.

[7] 峯松他，分析合成技術を利用した自己聴取音の作成に関する実験的検討, 日本音響学会春季講演論文集, 1-7-8, 211-212, 2000.

[8] 森他，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, 127, 8, 1268–1269, 2007.

[9] 森田他，気導音と骨導音を考慮した音声合成モデルの検討, 日本音響学会春季講演論文集, 1-R-9, 597–600, 2019.

[10] S. Hansen, “Air conducted and body conducted sound produced by own voice,” *Canadian Acoustics*, 26, 2, 11–19, 1998.

[11] 中川，吉田，対話型進化計算手法を用いたフォント自動生成システム，情報処理学会研究報告，Vol.2010-HCI-136 No.3，2010.

[12] H. Takagi *et al.*, “Interactive evolutionary computation-based hearing aid fitting,” *IEEE Transactions on Evolutionary Computation*, 11, 3, 414–427, 2007.

[13] 佐野，補聴器の進歩と聴覚医学「補聴器の fitting について」，*Audiology Japan*, 60, 201–209, 2017.

[14] S. Ding *et al.*, “Golden speaker builder – An interactive tool for pronunciation training,” *Speech Communication*, 115, 51–66, 2019.

[15] Z. Yi *et al.*, “Voice Conversion Challenge 2020 – Intra-lingual semi-parallel and cross-lingual voice conversion –” *Proc. Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020*, 80–98, 2020.

[16] D. Kao *et al.*, “The effects of a self-similar avatar voice in educational games,” *Proc. ACM on Human-Computer Interaction*, 5, CHI PLAY, 238, 1–28, 2021.