

GA を用いた頭外音像定位音の前後誤判定の改善

和田 安弘 南部 功夫 ○佐藤 祐哉 (長岡技術科学大学)

Improvement of front-back confusion on Out-of head sound localization sound using Genetic Algorithm

*Y. Sato I. Nambu Y. Wada (Nagaoka University of technology)

Abstract— Out-of-head sound image localization technology is a technique to obtain a stereoscopic sound image even at listening to earphone by convolving characteristics called head related transfer function (HRTF) with the sound source. In this technique, although the sound image is generated in the front, a misjudgment that the subject perceives the sound image behind the subject is a problem before and after the sound image. In this research, focusing on band information called spectral cue for this misjudgment, HRTF characteristics convoluted with sound source for individual characteristics of each subject are compared with genetic algorithm which is a point search algorithm established To reduce the incidence of erroneous judgments. As a result, it was confirmed that the decline rate of the subject as a whole decreased significantly before and after GA adaptation of the subject as a whole, and the method proposed this time was effective.

Key Words: 遺伝的アルゴリズム, 頭外音像定位, 仮想音, 頭部伝達関数, 前後誤判定

1 始めに

ヒトは音源から鼓膜位置までの空間伝達特性や耳介形状、頭部形状に影響を受けた音を聞いている。音に影響を与える空間の伝達特性を表したものを頭部伝達関数 (HRTF : Head-Related Transfer Function) と呼び、HRTF を音源に畳み込む事で、イヤホン・ヘッドホン受聴でも、実音源のような立体的な空間音像を再現する、頭外音像定位技術がある³⁾。この技術を用いて、仮想聴覚ディスプレイ (VAD : Virtual Sound Display) という、仮想的に立体的な音場を再現する研究が盛んに行われており¹⁾、すでに一部では聴覚障害を負った人へのリハビリ等に実応用されている⁴⁾。だが、この技術には、本来音像を定位させたい位置と実際に被験者が知覚する位置がずれが生じる事があり、特に、前方向に定位させたい音像が後方向に定位してしまう前後誤判定という問題を持つ。

なぜこのような音像の誤知覚が起こるのか、ヒトが音像をどのように定位させているかについては、これまで広範に研究されてきた³⁾⁴⁾。音像定位の主要要因は両耳間での音の到来する時間差 (ITD : Interaural Time Difference) や音圧差 (ILD : Interaural Level Difference) である「二重理論」を基に音源の方向、距離を推定していると考えられている。しかし、前後誤判定においてこれらの手掛かりは有効であるとは言い難い。なぜなら、もしこのバイノーラルキューが前後誤判定の主要な手掛かりであるならば、両耳間差がほぼ存在しない正中面において、前後誤判定が発生することはないからである。

ヒトの上下方向の音像定位について、特に正中面定位に関する多くの研究がなされている⁶⁾。正中面では、頭部形状の対称性から両耳間差はほとんど考慮する必要はない。しかし、正中面上における前後・上下方向の音像定位は、水平面と同様に定位可能であること、つまり、前後・上下方向の定位は、両耳間差が得られなくとも可能となることを示唆している。この結果から前後・上下定位に関して周波数領域から得られる周波数スペクトルの手掛かりである「スペクトルキュー」を用いて定位を実現していると考えられている³⁾。

また、HRTF はヒトの頭部、耳介形状に影響され、個人毎にその特性を大きく変える (Fig.1 参照)。各個人の理想的な伝達特性を再現できれば正確な音源定位が実現できるが、個人に理想的な特性を得るには、無響室のような特殊な測定環境、長時間の被験者拘束、高価な測定機器、仮想音への慣れ等様々な要素が関係してくる。また、この測定環境を整えて測定した特性でも被験者に対して必ずしも定位感が良く聞こえる保証があるわけではない事が知られている⁵⁾。そこで、データベースの HRTF を用いて被験者の HRTF を取得する研究がなされている。遺伝的アルゴリズム (IGA : Interactive Genetic Algorithm) を使用した研究¹⁰⁾では、HRTF の全帯域をモデル化し、そのモデル化係数を調整する事で、既知方向の HRTF から未知方向の HRTF を被験者の主観評価を用いて算出できる事を示唆された。そこで、本研究では、被験者の主観評価を用いた GA を用いて、HRTF における前後誤判定の原因と推定されるスペクトルキューを調整する事によって、前後誤判定の発生率を抑制する手法として提案する。

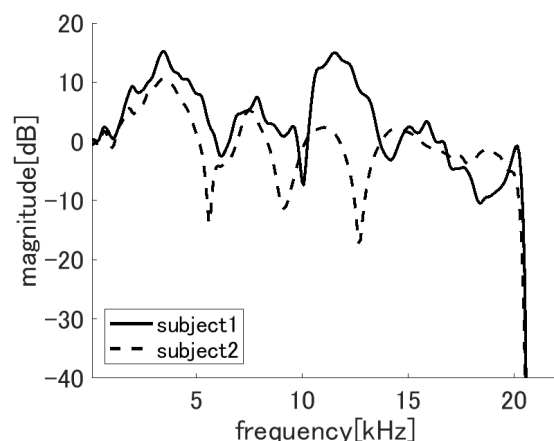


Fig. 1: difference of HRTF among humans

2 提案手法

2.1 前後知覚に影響を持つスペクトルキュー

本研究で使用したスペクトルキューについて、特に前後誤判定の手掛かりとなりうるスペクトルキューを先行研究の知見を基に選定した。以下に、参考にした論文の一部を記載した。測定した HRTF と測定した HRTF の内、5.7~11.3kHz または、8~16kHz を欠落させた HRTF とを比較した結果、その定位誤差・前後誤判定率の間に有意差が確認された⁷⁾。HRTF の 4kHz から始め、最も近いノッチ、ピークをそれぞれ、 $N_1, N_2, \dots, N_i, P_1, P_2, \dots, P_i$ と定義した時、 N_1 (4kHz 近傍)、 N_2 (7~9kHz 近傍) の帯域の有無で前後誤判断の発生率が劇的に改善されたという報告がある⁸⁾。また、ヒトの前方知覚、後方知覚はそれぞれ 4~8kHz のノッチと 13kHz 以上のピーク、10~12kHz のピーク、ノッチであることも報告した⁹⁾。

2.2 頭部伝達関数のモデル化

一般的にモデル化を行うにあたり多くのデータを使用すればするほどモデルはより汎化なものとなる。インターネット上で公開されている RIEC の頭部インパルス応答 (Head-Related Impulse Responses; HRIR) のデータベースを使用した。HRIR とは HRTF に逆フーリエ変換を行った時間領域の信号である。RIEC は東北大学が公開しているデータベースであり、105 名分のデータが開示されている。開示されている HRIR のデータ形式は、仰角、方位角それぞれ 5 度間隔、サンプル数: 512、サンプリング周波数: 48kHz となっている。本研究では、仰角 0 度、方位角 15 度間隔 24 方向の、 $24 \times 512 \times 105$ データを利用した。

前処理として、サンプリング周波数を 48kHz から 44.1kHz にダウンサンプリングし、フーリエ変換により HRIR から HRTF を生成、2.1 節を基に、HRTF を 3~13kHz で帯域制限した。その後、データセットを主成分分析にかけ、HRTF の共通成分である一般基底関数 C と各被験者個人の特性を表す重み係数 ω に HRTF を分離、主成分の寄与率 99.9% を達する重み係数 30 個を使用した。

2.3 遺伝的アルゴリズム

2.2 節にて、HRTF をモデル化した重み係数に用いて、被験者に適した HRTF 特性を取得する。その手法として、被験者評価を用いた対話型遺伝的アルゴリズム (IGA : Interactive Genetic algorithm)¹⁰⁾ を採用した。GA とは、生物が環境に適応する際の遺伝学的変化の諸概念たる染色体の交配・突然変異・自然淘汰を問題解決方法に見立て、目的変数の解を探る手法である。このアルゴリズムの利点として、探索空間が広範であっても近似解への探索速度が速い点や局所解からの脱却手法がアルゴリズム自体に内在している点が挙げられる。アルゴリズムのパラメータはシミュレーションを通して決定され、Table1 に設計パラメータを記載した。

Table 1: Design parameters of GA

parameter	value	unit
Crossover rate	80	%
Mutation rate	1	%
Individuals	5	
Gene length	16	bit
Contribution rate	99.9	%

3 被験者実験

3.1 被験者応答

2.1 から、被験者本人の HRTF と被験者が聞いた仮想音生成に使用した HRTF における、前後知覚に関係する部分帯域のスペクトル歪みが大きければ大きいほど、前後誤判定に大きくなるという仮定を基に、HRTF データベースの中から被験者の測定した HRTF との SD が最大となる HRTF を選択し、その HRTF を使用した仮想音、及び GA 適応後の HRTF により作成した仮想音での定位試験を行った。定位試験では、被験者には Fig.2 に記載した MATLAB GUI を提示した。

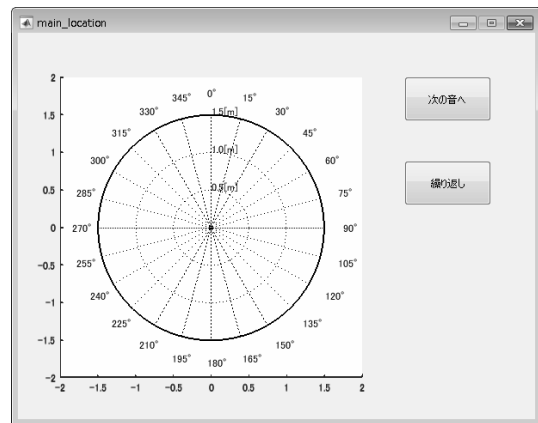


Fig. 2: A GUI is presented in which the center position is the subject position, and the direction and angle from the subject position are shown. The subject listens to the presented virtual sound and shows the perceived position by subjective evaluation

3.2 測定項目

3.2.1 前後誤判定

前後誤判定は前方位置に提示した仮想音が両耳軸を対称に後方位置に定位する、または、その逆の現象をさす。例えば、今回の実験で使用する 6 方向に関して説明すると、仮想音の提示方向が 0° に対して、被験者知覚が 180° である時、同様に提示方向が 60° に対して、被験者知覚が 120° のような仮想音の誤知覚を示すものである。今回は 6 方向を実験で使用するため、その対称位置から $\pm 30^\circ$ の範囲を前後誤判定の範囲とした。つまり、60° を提示方向とした時、被験者の知覚位置が 90° から 150° であった時前後誤判定として記録した。また、前後誤判定率とは式 1 に示したように、仮想音の総提示数に対して前後誤判定が発生した回数を百分率で表したものである。

$$e_{FBC} = \frac{\text{No. of front back confusion}}{\text{No. of total response}} \cdot 100(\%) \quad (1)$$

3.2.2 定位誤差

定位誤差は式2に従う。添え字*i*は実験全体での被験者に対する仮想音の提示回数であり、*T*が提示した仮想音の提示角度 (Target angle)、*P*が被験者が知覚した角度 (Perceived angle)を示す。*e*は提示した仮想音の提示角度と被験者が知覚した角度差を示す。例えば、提示角度が60°であり、被験者の知覚位置が45°であったならば、その定位誤差は15°となる。

$$e_i = |T_i - P_i| \quad (2)$$

4 実験結果

4.1 前後誤判定率

被験者全体の結果を確認した。GA 適応前後における各被験者毎に算出された前後誤判定率の平均値を用いて、1 標本 t 検定を行った結果、GA 適応前後で有意水準 5% で有意差を観測した (p 値: p = 0.0207)。つまり、今回の提案手法である、HRTF の部分帯域を GA を用いて被験者に適した特性を生成する手法が前後誤判定の発生率の改善に有効であることを確認した。

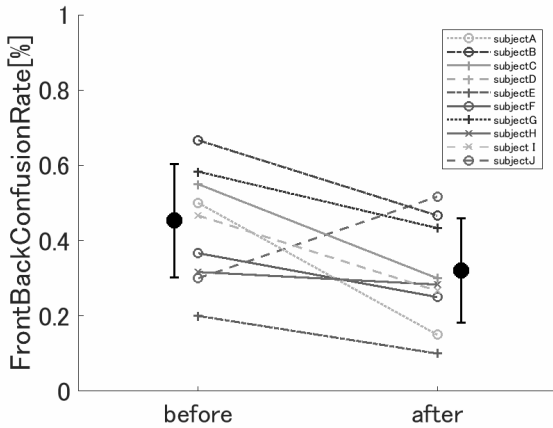


Fig. 3: result of front-back confusion rate

次に被験者個人個人の結果を確認する。table2は実験結果を被験者毎に集計したものである。beforeがGA 適応前、すなわち初期値における前後誤判定の発生率を表し、afterはGA 適応後の前後誤判定の発生率を表す。Improvement rateの行が黒く太字になっているものが、前後誤判定率がGAの適応前後で改善した被験者を表し、それ以外のものがGAを適応した結果、発生率が増加してしまった被験者を表す。また、significant differenceは1標本t検定を行った検定結果を表す。*が有意水準5%、**が有意水準1%、n.s.が非有意を表す。被験者10名中9名の被験者で前後誤判定率が改善し、そのうち4名の被験者がGAの適応後有意に前後誤判定の発生率が減少した事、被験者1名が有意に悪化した事を示している。

Table 2: result of front-back confusion rate

subject	A	B	C	D	E	F	G	H	I	J
before[%]	50.0	66.7	55.0	58.3	20.0	35.0	58.3	31.7	46.7	30.0
after[%]	15.0	46.7	30.0	43.3	10.0	26.7	43.3	28.3	26.7	51.7
Improvement rate[%]	35.0	20.0	25.0	15.0	10.0	8.3	15.0	3.3	20.0	-21.7
significant difference	**	**	n.s.	n.s.	n.s.	n.s.	*	n.s.	*	*

4.2 定位誤差

被験者全体の傾向を確認する。Fig.3は、全被験者の定位誤差の集計結果である。図の参照については、Fig.2と同様である。GA 適応前後における各被験者毎に算出された定位誤差の平均値を用いて、1 標本 t 検定を行った結果、GA 適応前後において有意水準 5% で有意差が確認されなかった (p 値: p = 0.4372)。これは、今回の提案手法が定位誤差に影響を与えない事を示している。

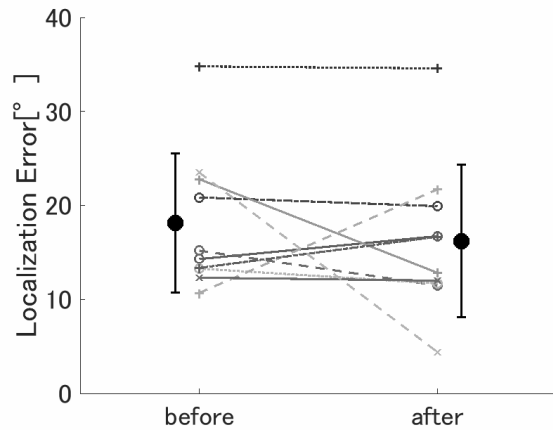


Fig. 4: result of localization error

次に、被験者個人個人の結果を確認する。table3は実験結果を被験者毎に集計したものである。表の参照の仕方は節の2と同様である為割愛する。被験者10名中8名の被験者で前後誤判定率が改善し、そのうち1名の被験者がGAの適応後有意に定位誤差が減少した事、被験者1名が有意に悪化した事を示している。

Table 3: result of localization error

subject	A	B	C	D	E	F	G	H	I	J
before[°]	41.9	28.2	34.3	24.5	20.2	27.3	27.9	34.3	24.2	32.6
after[°]	41.3	28.3	37.1	26.5	23.0	25.2	33.3	37.3	24.3	33.4
Improvement rate[°]	-0.7	0.1	2.8	2.0	2.8	-2.1	5.3	3.0	0.1	0.8
significant difference	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	**

5 考察

5.1 提案手法の評価

本提案手法を用いた被験者実験において、9名の被験者で提示した仮想音の定位感評価を行うことで前後誤判定率を改善することが出来た。被験者は平均的に10回前後の重み係数の更新によって、実際のHRTFを測定する事無く前後誤判定の改善できることを示した。被験者のタスクは、仮想音を2つ提示し相対評価のみを行ってもらうのみである。その為、仮想音に慣れていない被験者においても比較的、少なくとも仮想音を絶対評価するよりも簡便に評価することが可能であると考えられる。その為、実験結果を鑑みて、すべての被験者に使用できる汎用的な処理というにはまだまだ改善点も多いが、前後誤判定を改善する手法としては有効

であると考え。

また、頭外音像定位技術には、前方向の定位と後方向の定位を間違える前後誤判定と同様に、上方向の定位と下方向の定位を間違える上下誤判定 (Up-Down Confusion) が存在する。本研究では、周波数帯域の調整によって前者の問題を改善できる結果を得た。その為、同様に後者の問題にも適応できるのではないかと考える。

5.2 有意差を確認した被験者と確認できない被験者の差

前後誤判定率が改善方向に向かった被験者 9 名の内、有意差が確認された被験者 4 名 (被験者 A,B,G,I) と有意差を確認できなかった被験者 5 名の差を調べた。

Fig.4 は被験者 A、つまり、有意差を確認できた被験者の各方向毎の前後誤判定の発生率をまとめたものである。before が GA 適応前、after が GA 適応前であるが、その前後で前後誤判定率が全体的に減少している傾向にある。その一方で、Fig.5 は被験者 C、つまり、有意差を確認できなかった被験者の各方向毎の前後誤判定の発生率をまとめたものである。GA 適応前後で前後誤判定の発生率が変化していないことが分かる。特に 0 方向と 120 方向が顕著である。このような傾向が有意差を確認できなかった被験者に見られた。これは、初期音から GA を更新してもより良い個体が生成されず、初期値で GA を更新し続けたために、収束が急速に進んでしまい、個体の多様性が失われた可能性が挙げられる。今回採用した IGA は被験者の評価値を基にデータベースの HRTF を参考に設定した初期値を目的方向の被験者の HRTF に適した特性を生成するように学習していく。そのため、初期値から変化が起こらないとき、GA の更新が繰り返されると、似たようなデータ同士で次の個体を生成してしまうことがある。今回の結果はその一例であると考えられる。GA では突然変異というパラメータを使用し、できる限り局所解から脱却できるように構成されているが、有意差を確認できなかった被験者にはうまく機能しなかった可能性がある。

Table 4: front-back confusion rate each direction in subject A

	average[°]	0°	60°	120°	180°	240°	300°
before	50.00	100	80	20	0	0	100
after	15.00	50	0	30	0	0	10

Table 5: front-back confusion rate each direction in subject C

	average[°]	0°	60°	120°	180°	240°	300°
before	31.667	80	0	70	0	30	10
after	28.333	80	10	70	0	0	10

6 結論

本研究では、頭外音像定位技術の前後誤判定改善のために GA を用いて部分帯域を最適化する主観適応方法を提案した。

データベースの HRTF の内、前後知覚に関わる帯域を主成分分析にかけ、データベースの HRTF 共通成分たる一般基底関数と被験者毎の固有成分たる重み係数

を生成、重み係数の平均値、標準偏差から初期値を生成する。被験者の主観評価を基に重み係数値を更新し、前後誤判定が発生しない被験者に適した HRTF を取得する。

被験者試験を通して提案手法によって主観適応し、生成した HRTF の前後誤判定、定位誤差の検証を被験者 10 名を募って実験した。この結果、被験者全体として、前後誤判定率に関して 1 標本 t 検定を行った結果、初期値となった他人の HRTF に対して、有意水準 5% で有意差を確認し、また、被験者個別に見ると 10 名中 4 名の被験者で同様に有意差を確認した。しかし、今回の提案手法でも前後誤判定率が変わっていない被験者や逆に改悪してしまった被験者も同様に確認されており、すべての被験者に使用できるような汎用性のある処理には至っていない為、処理の追加、最適化手法の改善等改良の余地はまだ存在する。以上の結果から主観評価を用いた GA を用いて部分帯域を最適化する手法は前後誤判定に対して有効である事が示唆された。

6.1 参考文献

参考文献

- 1) Begault, Durand R and Wenzel, Elizabeth M : Techniques and applications for binaural sound manipulation, The International Journal of Aviation Psychology, 2 巻 1 号, 1/22 , (2003)
- 2) Blauert, Jens : Spatial hearing: the psychophysics of human sound localization, 日本音響学会誌, 73 巻 3 号, MIT press , (1997)
- 3) 長谷川 侑太: 粒子群最適化を用いた頭部伝達関数の主観的応に関する研究, 長岡技術科学大学 修士論文, (2016)
- 4) 岩谷幸雄 : 頭部伝達関数による音像定位, 日本音響学会誌, 73 巻 3 号, 173/180 (2017)
- 5) Zotkin, DYN and Hwang, Jane and Duraiswaini, R and Davis, Larry S : HRTF personalization using anthropometric measurements, Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on, leee , (2003)
- 6) J. Blauert, "Sound localization in the median plane," Acustica, 22 巻, 205/213 (1969)
- 7) Langendijk, Erno HA and Bronkhorst, Adelbert W : Contribution of spectral cues to human sound localization, The Journal of the Acoustical Society of America, 112 巻 4 号, 1583/1596 (2002)
- 8) Iida, K., Itoh, M., Itagaki, A., and Morimoto, M : Median plane localization using a parametric model of the head-related transfer function based on spectral cues, Applied Acoustics, 68 巻 8 号, 835/850 (2007)
- 9) Iida, K. : Median plane localization using a parametric 音講論 (秋), 473/476 (2010)
- 10) Cheung, N.-M., Trautmann, S., and Horner, A. : Head-related transfer function modeling in 3-D sound systems with genetic algorithms , in Acoustics, Speech and Signal Processing, 6 巻, 3529/3532 (1998)