

# 頭部伝達関数の機械学習に基づく 前後方向の音像識別に関する検討

瀬音 光一（永田研究室）

## 1 はじめに

前後方向の音像の識別には、頭の形、特に耳介の形と前後の非対称性が大きく影響していることが知られている。この非対称により、同一の信号音でも前方から来た場合と後方から来た場合とでは、各々の周波数特性が異なる。この周波数特性上の方向定位に関する手掛かりをスペクトルキューと呼び、前後方向のスペクトルキューについては、頭部伝達関数（HRTF：Head-Related Transfer Function）の4kHz付近に存在するピーク（P1）、及び4kHz以上に存在する最初のノッチ（N1）とその次のノッチ（N2）が知られているが[1]、HRTFは音源の信号が既知でないと推定できないはずであり、またスペクトル上のノッチは背景雑音によって容易に消失してしまうため、ヒトがノイズの存在する環境の中でその差異をどのように識別し、学習しているかは明らかではない。そこで、本研究では特に音像制御の難度が高いとされている前後の定位[2]について機械学習を用いてその学習が可能であるかを検討する。

## 2 断片的な信号音による到来方向識別

音響伝達関数推定に使われる掃引正弦波や白色雑音といった人工的な信号を方向識別の学習信号として用いれば、前後の識別は比較的容易であると考えられるが、ヒトは測定用信号ではなく、時間的にも周波数の上でも断片的で片寄った信号から到来方向に関する学習を行っているものと考えられる。また先述したように、HRTFは音源の信号が既知でないと推定できないはずであり、背景雑音が伝達関数上のノッチを隠し、さらに学習の精度を低下させるものと考えられる。

そこで、本研究では伝達関数測定用の人工的な信号ではなく、実世界で到来するような断片的な信号から不完全な前後のHRTFと両耳間レベル差（ILD：Interaural Level Difference）の集合を求め、これらを入力とした機械学習による前後の識別について検討する。

### 2.1 前後の頭部伝達関数の測定

HRTF推定に使用した音源は、音声、効果音、ボーカル入り音楽の3種類で時間長は各々3秒弱である。音源の内訳は、音声72、効果音60、音楽88、合計220であり、録音はダミーヘッドの両耳にマイクロホンを着装し、図1に示す環境で行った。また、録音信号には別途ダミーヘッドで録音した左右の背景雑音をSN比10dBとなる

ように加えた。背景雑音を加えた録音信号および、音源信号に対し1024点のDFTスペクトルからクロススペクトル法によりHRTFを算出した。

上記の環境で、白色雑音と220の信号からHRTFを求めた。白色雑音のHRTFを図2に、220の信号ごとに得られたHRTFのうち例として3つを図3に示す。図2と図3を比べると、白色雑音では、6kHz付近のノッチで前後の判断が可能であり先行研究[1]と一致する結果が得られた。一方、雑音重畳両耳録音信号からは前後の特徴を見出すことが困難であった。

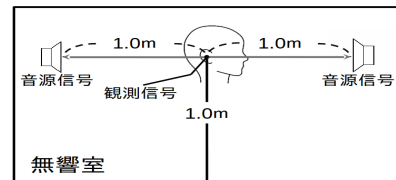


図1: 録音環境

### 2.2 前後の両耳間レベル差の測定

ILDもHRTFと同様に、左耳の録音信号と右耳の録音信号から1024点のDFTによるクロススペクトル法を用いて算出した。白色雑音の両耳間レベル差を図4に、220の信号ごとに得られた両耳間レベル差のうち例として3つを図5に示す。図4と図5を比べるとどちらも10kHz付近で前後の差異が確認できる。

## 3 識別テスト

### 3.1 識別結果および考察

HRTFによる識別では、前後の差異が目視でも確認できた4~12kHzの帯域を使用し、ILDの場合では8~15kHzの帯域を使用した。次元数は各々186点、163点であり、教師データ220、評価データ110、テストデータ110で機械学習を行った。機械学習には、中間層1層のニューラルネットワークを使用し、ラベルは音源が前方にある場合のHRTFとILDに0を、後方にある場合のHRTFとILDには1を付与している。識別実験の結果を表1に示す。

この結果からHRTFを用いた場合は前後の識別が困難であり、これは左右のHRTFどちらも同じ結果となった。対して、ILDによる識別の場合は識別率が最大95.5%となり、教師データが220と比較的少ない現段階では十分識別できたといえる。

	HRTF(左)	ILD
テストデータの識別率	53.5	94.1
評価データの識別率	56.8	96.9
教師データのみの識別率	56.7	100
学習回数	50	15
中間層のノード数	50	15

表1: HRTFとILDの前後識別結果

## 4 むすび

本研究では、雑音下での断片的な信号からのHRTFとILDの機械学習を用いた前後方向の識別を行った。その結果、到来音そのものよりも有利なHRTFの場合でも識別率53.5%と識別が困難であり、対して聴覚において両耳信号から計算可能なILDの場合は識別率が94.1%と高い識別性能であった。このことから、断片的で片寄った信号や背景雑音を含んだ信号など実世界の到来音においては従来の研究で主張されているHRTFよりもILDによる前後の識別の方が容易である事が分かった。

学習器で識別できるということは、実世界においてヒトは両耳間の違いによって前後の差異を学習しているという仮説が成り立つといえる。長年、ヘッドホン受聴による音像制御が研究されているが、特に前後の正面定位に関しては十分な性能のシステムが実現されているとはいえない。本研究の結果は音像制御において左右差を強調する方が前後の定位には有効かもしれないという可能性を示した。

## 参考文献

- [1] K. Iida et al., Applied Acoustics, vol.68, pp.835-850, 2007.
- [2] 飯田一博、森本政之、「音響サイエンスシリーズ2 空間音響学」日本音響学会編、コロナ社、2010

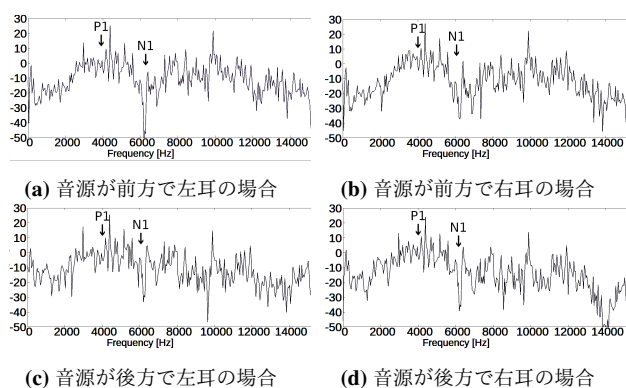


図2: 白色雑音を用いて測定した頭部伝達関数

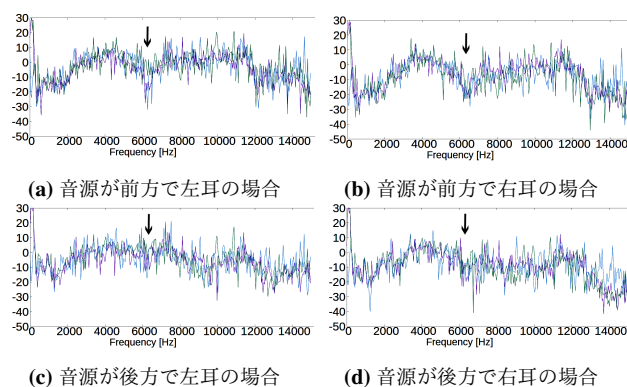


図3: 雑用重畳録音信号から測定した頭部伝達関数

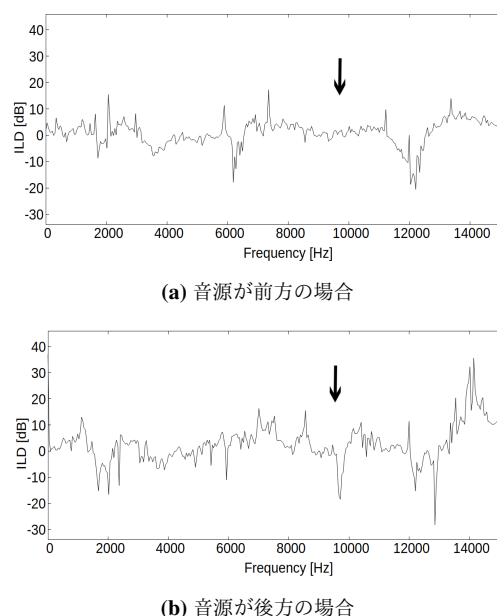


図4: 白色雑音を用いて測定した両耳間レベル差

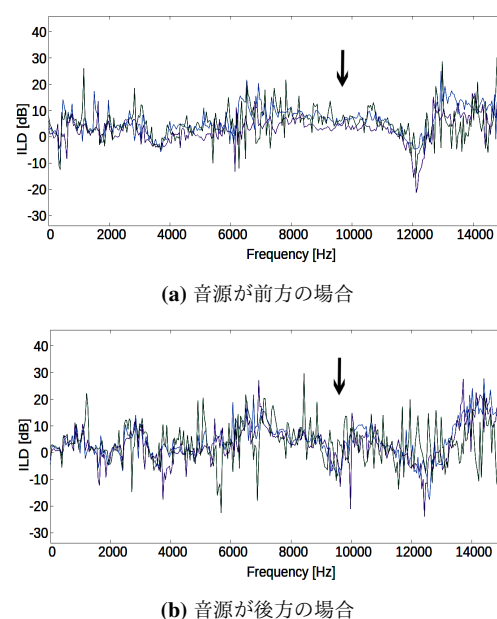


図5: 雑用重畳録音信号から測定した両耳間レベル差の例