

両耳神経モデルの考察とその音場解析への応用

Considerations on binaural model and its application to analyses of sound fields

- | | | |
|-------|---|-------|
| 代表研究者 | 東京電機大学工学部助手
Instructor, Fac. of Engineering, Tokyo Denki Univ.
Kaoru OKABE | 岡部 馨 |
| 協同研究者 | 東京電機大学工学部教授
Prof., Fac. of Engineering, Tokyo Denki Univ.
Tanetoshi MIURA | 三浦 種敏 |
| | 音響工学研究所取締役社長
President, Acoustic Engineering Laboratory
Masaru KOYASU | 子安 勝 |
| | 日東紡音響エンジニアリング取締役社長
President, Nittobo Acoustic Engineering
Toshiaki SHIGETA | 茂田 敏明 |
| | 東京医科歯科大学難治疾患研究所神経生理部門教授
Prof., Dept. of Neurophysiology, Medical Research Institute,
Tokyo Medical and Dental Univ.
Keiichi MURATA | 村田 計一 |
| | 富山大学工学部教授
Prof., Fac. of Engineering, Toyama Univ.
Takeshi NAKAYAMA | 中山 剛 |

Recent study proposes the binaural model in which the functional role of superior olivary complex are taken into account. Running interaural cross-correlation function simulates the behavior of the medial superior olivary complex (MSO) cell, which acts as the interaural time difference (ITD) detector and interaural coherency displayer. The lateral superior olivary complex (LSO) displayer illustrates the short-term interaural level difference and binaural energy sum. The contra lateral inhibition is introduced in LSO processing.

The model was applied to analyze the subjective evaluation of concert hall sound fields. The source was classic orchestra music and 24 subjects assessed the goodness of the hall acoustics in 7 categories. Factor analysis indicated four factors were sufficient to explain 85% of total data variance. The most important factor was subjective spaciousness. Multiple regression analysis found out that the standard deviation (SD) of the ITD around 1 kHz and SD of the ILD around 4 kHz were strongly correlated to spaciousness.

研究目的

今日、都市交通騒音・労働環境騒音などの公害的音響からコンサートホールや録音スタジオなど芸術的音響まで、我々をとりまく音環境に対する

分析・評価法確立への要求は強く、かつ多岐にわたっている。しかし現時点では、計測マイクロホンでとらえ得る音場の物理現象と最終的な人間の主観評価の間には大きな隔たりが存在し、まだ両

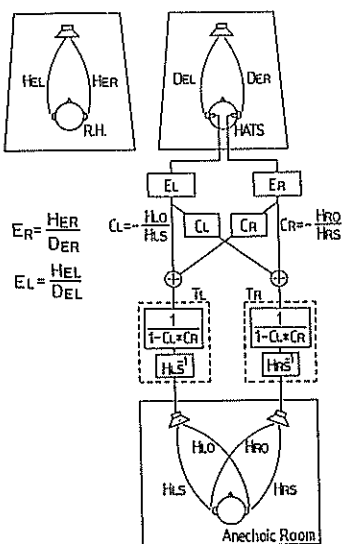


図1. HATS を用いた音場再生系 Orthostereophonic System (OSS)

者の定量的関係を説明できる段階とはいえない。その理由のひとつに、人間は左右の2つの耳で音響情報を捕らえ、それらを耳から脳までの神経ネットワークで総合処理し、情報抽出をおこなっている点が挙げられる。その結果、人間は音により環境を認識したり、周囲騒音の中から音声などの必要な情報を選択聴取することが可能になるが、従来の音場計測では単純なマイクロホンシステムの出力のみしか評価されておらず情報抽出が不十分なため、最終的な主観評価と一致しない結果が生じると考えられる。そこで本研究では、人間が音を聴取するという立場から音場分析・評価の問題にアプローチするため、人間の胴体、頭部、外耳を模した Head and torso simulator (HATS) の両耳マイクロホンで音場情報を収集し、これらを内耳以降の両耳神経ネットワークを模擬したモデルに従い処理し、音場知覚との因果関係を調べ、音場分析・評価法の基礎確立をめざす。その一つの応用としてコンサートホール音場評価を試みた。

研究経過および成果

1. コンサート音場の主観評価実験

上述のような問題に合目的的にアプローチする

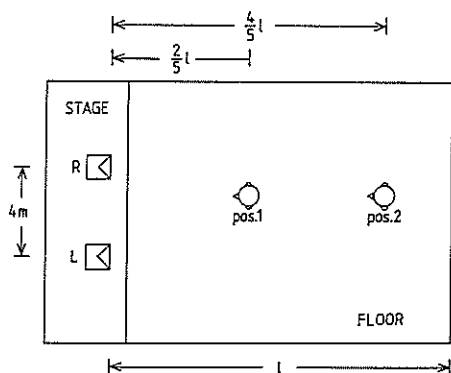


図2. ホール音場における收音条件. 收音位置はホールの寸法比から決定した。

ためには、HATS で收音した両耳信号を受聴者の両耳に正確に再生するシステムが必要となる。なぜなら、このようなシステムが実現できれば、さまざまな音場に受聴者がいちいち出向いて、音を聴取しその主観評価を求める必要がなくなると同時に、HATS で各種音場を收音しておけば、任意の時間・場所で主観評価実験を精度よく繰り返せ、異なる音場の比較が容易になるためである。

その実現法の一つに、三浦らの提案した Orthostereophonic System (OSS) がある。OSS では、図1に示すような二つのスピーカから HATS 両耳信号を再生するが、よく知られているように左スピーカから右耳、右スピーカから左耳へ回り込むクロストークが存在するため、HATS の両耳信号が受聴者の両耳に正確に再現できない。このため、図中に示した OSS ネットワークと称する前置デジタルフィルタを用い、このクロストークをキャンセルしている。その結果、HATS 両耳信号を正しく受聴者の両耳に再生することが可能になり、受聴者はあたかも原音場にいるのと同等の受聴が可能になる。

音場評価実験の目的は、ホールの違い、座席位置の違いから生じる音質の善し悪しを定量的に表すことにある。その後で、これらが両耳モデルから抽出された物理特性とどのような因果関係を持っているかを調べる。具体的な実験ではステージ上に2台の無指向性スピーカを設置し、無響室録音の音楽、いわゆるドライ・ミュージックを再

表 1. ホール音質評価のための 7 段階カテゴリ.

非常によい	6
よい	5
少しよい	4
中間	3
少し悪い	2
悪い	1
非常に悪い	0

生し HATS で録音した。各ホールで、同一のドライ・ミュージックを同一のスピーカシステムを用い再生すれば、その音質の違いは純粋にホールの音響特性の違いのみに起因するものと考えられる。

收音条件を図 2 に示すが、收音位置はホールの寸法比から求めた。対象としたホールは、200 から 2000 席程度の座席数を有するもの 11 カ所で、演奏会専用ホールや多目的ホールがほぼ半数ずつ含まれている。

音楽ソースには、モーツァルト作曲オペラ「フィガロの結婚」序曲（「無響室のオーケストラ」：日本コロムビア CD）から広帯域信号を含む約 20 秒を選択した。收音に際しては、スピーカに入力するパワーはホールの大小には関わらず一定とした。

また、收音位置は原則として各ホール 2 ポイントずつとしたが、主観評価実験の際に受聴者の負担を減らすため、予備実験で大きな差の認められなかったホールは 1 ポイントのみとし、総数 18 ポイントで最終評価実験を行った。主観評価実験

表 2. 音質評価の際に受聴者に提示した表現語。受聴者は選出の際には適切なものを任意個選択してよいと指示されている。

派手な	おとなしい
明るい	暗い
輝きのある	くすんだ
さわやかな	生気のない
歯切れのよい	モガモガした
きめの細かい	きめの粗い
柔らかい	硬い
澄んだ	濁った
のびのびした	つまつた
潤いのある	力サカサした
落ち着いた	騒々しい
豊かな	貧弱な
どっしりした	軽々しい
迫力のある	迫力のない
拡がりのある	広がりのない
なめらかな	ざらざらした
バランスのとれた	アンバランスな
素直な	くせのある
遠く感じる	近く感じる
ヌケのよい	こもった
奥行きのある	奥行きのない

では、18 種類の音刺激を OSS でランダムに再生し、表 1 の 7 段階カテゴリ表現を用いて評価した。数回の練習の後これを繰り返し、その平均値を評価値として採用した。受聴者は正常聴力を有する成人男女 24 名で、録音エンジニア、工学専攻および音楽専攻の学生から成る。

こうして求めた 18×24 の評価値を、中山らの因子分析を用いた音質評価法により解析し、心理評価空間における各ホールの位置、各受聴者の重み付けを求めた。また、後の解析結果の解釈の手助けとするため、受聴者から評価値を求める際に表 2 の表現語から適切なものを任意個選出させた。

因子分析の結果、ほぼ 4 次元空間で全データ変動の 85% が説明できた。特に、第一、第二因子で 60% 以上のデータ変動が説明できた。そこで図 3 には第一、第二因子に基づく心理空間を、図 4 には同空間に対する各受聴者の重みづけベクトルを示す。図 3 の A~K はホールを、添え字の数字は座席の位置を表す。

表現語の出現頻度と心理空間との相関を求めた結果、第一因子は広がりや奥行きなど、音場の空間性を表す因子と考えられた。一方、第二因子は

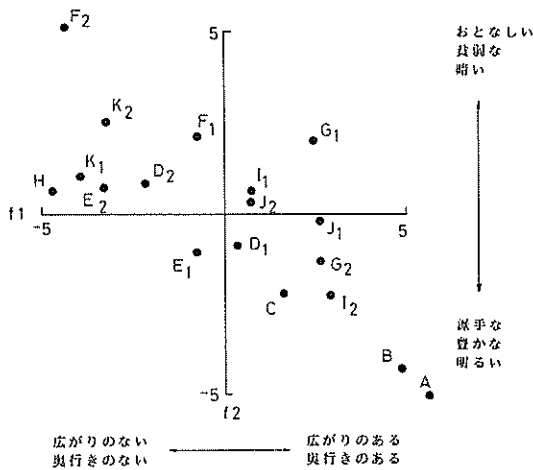


図3. 2次元心理空間における各ホールの位置。A～Kはホールを表し、添え字は座席位置を表す。

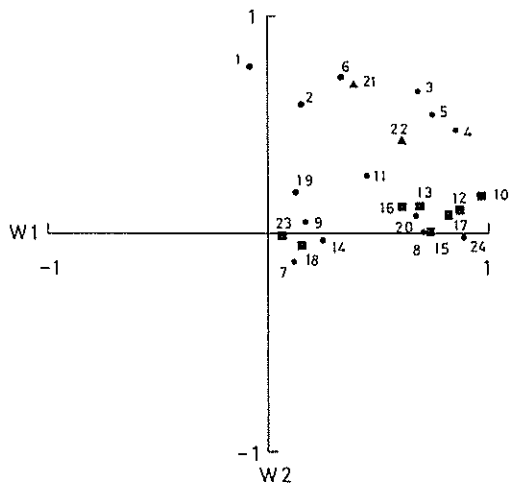


図4. 各受聴者の重みづけベクトル。四角は録音エンジニアを、三角は音楽専攻学生を、丸は工学専攻者を表す。

「派手な」、「豊かな」、「明るい」など音色に関わる因子と推察された。また、図4の重みベクトルから、第一因子、すなわち音場の空間性を重視する受聴者が多いことが分かる。

2. 両耳神経モデルの導出

受聴者の両耳に入射した音響信号は、外耳道、鼓膜、耳小骨連鎖を経て内耳に至る。内耳では、音響-機械振動は電気的神経信号に変換される

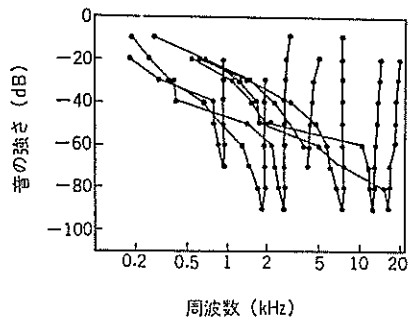


図5. 内耳蝸牛神経の周波数同調曲線(勝木らによる)。

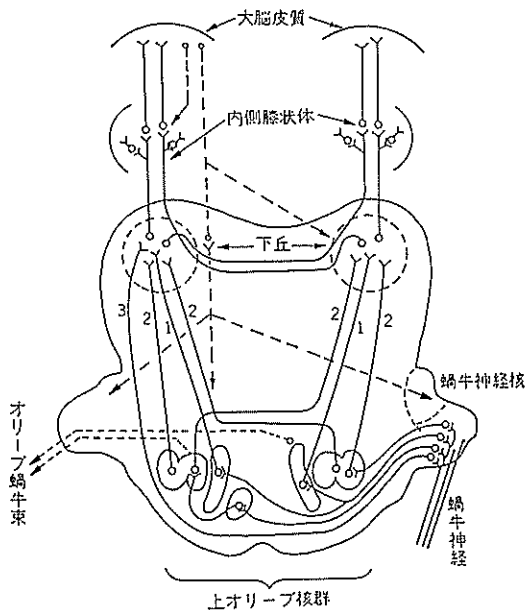


図6. 聴覚神経ネットワーク。

が、その際に図5に示すような周波数選択性をもつ狭帯域フィルタ処理が施される。

内耳から発生した神経信号は、図6に示す神経ネットワークを経て大脳皮質に到達し、最終的な知覚を生じる。特に、音源の方向探査や音場の空間性の知覚においては、人間が二つの耳を有していることが重要な役割を果たしていると考えられており、両耳に入射した信号がこの神経ネットワーク内でどのように処理されていくかを考慮する必要がある。

左右耳からの信号がはじめて交叉し、空間的情

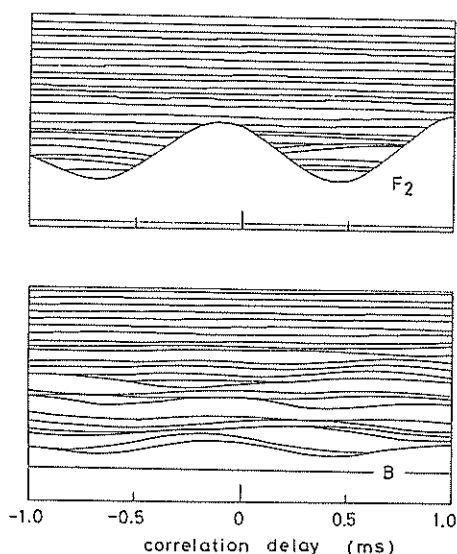


図7. MSOにおける処理を模擬するためのRunning interaural cross-correlation functionの例(ホールS2とB). 中心周波数は1 kHzで、直接波から150 msまで5 msごとに求めた。

報抽出が行われると思われる上オリブ核以降の処理が重要であるが、他方、大脳皮質に近づくにつれ情報処理が複雑になるうえに生理学的知見が限られるため、通常は機能的なモデルを仮定することが多い。

本研究では、上オリブ核の機能分類に基づく両耳神経モデルを検討し、前述の主観評価結果との関係を調べる。

上オリブ核はいくつかの部分に分類されており、最近の生理学分野でのデータによれば、両耳間時間差(ITD)に応じて反応し左右信号の相互相関演算にほぼ等しい振舞を示すMedial Superior Olivary Complex(MSO)と、両耳間振幅差(ILD)とITDの双方に反応するLateral Superior Olivary Complex(LSO)が機能上重要であると考えられている。特にMSOはILDにはほとんど影響を受けないため、ITDの検出器として動作すると考えられ、一方、LSOはILDに対してもよく反応し、反対耳からの強い抑制を受けるといわれる。一般に、反射の存在する空間で正しく音源位置を認識できるのは、最初に耳に入射した直接音

が後続の反射音を抑制する、いわゆる先行音効果のためといわれるが、LSOは先行音効果の生理学的根拠となっていると思われる。

そこで、まずMSOにおける信号処理を模擬するため、左右両耳信号のrunning interaural cross-correlation function(RIACC)を算出する。すなわち、 $r(t)$ 、 $l(t)$ をそれぞれ蝸牛で狭帯域フィルタ処理および半波整流された後の左右耳信号とすると、RIACC $\psi(t, \tau)$ は、

$$\psi(t, \tau) = \int_{-\infty}^t l(\theta + \tau) r(\theta) G(t - \theta) d\theta \quad (1)$$

で表される。 $G(t)$ は両耳の波形記憶能力が時間経過にもなると減衰する状態を表す重み関数で、

$$G(s) = \begin{cases} \exp(-s/T) & s \geq 0 \\ 0 & s < 0 \end{cases} \quad (2)$$

を用いる。ここで、 T は記憶減衰の時定数を表し、5 msとする。具体的な算出に際しては、M系列雑音を音源とし高速Hadamard変換を用いたインパルス応答計測法により求めたホールにおけるHATS両耳インパルス応答を利用する。次に、これに狭帯域フィルタ、半波整流処理を施した後(1)式を計算する。

図7はRIACCの算出例で、中心周波数1 kHzのデータを直接波から150 msまで表示した結果である。横軸は相関の遅延時間を表し、各5 msごとのRIACCを表示してあるため、各時刻におけるピークの高さと位置が相関度とITDの変動を表すことになる。この結果から、ホールF2では直接音の相関度が高く、かつ反射波の部分のピークがあまり顕著でないのに対し、ホールBでは全体的にピークの値が低く、反射波によりITDが大きく変化していることが分かる。こうしたRIACCでの違いは、ホールF2に比べホールBが広がりのある空間性を有すると判断された主観評価結果をよりよく説明していると考えられる。

次に、LSOの機能表示として短時間のILDを求めるが、その際には波形切り出し時間窓を定義する必要がある。一つには、先に示した指数関数的な時間窓を仮定することも可能ではあるが、ここではMooreらが心理実験により求めた人間の耳の時間窓特性(図8)を利用した。すなわち、

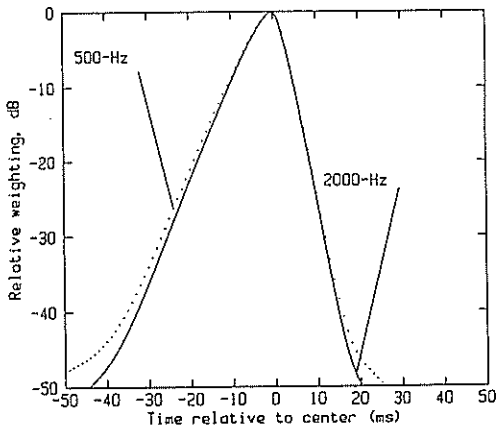


図 8. 人間の耳の時間窓特性 (Moore らによる).

$$ILD(t) = 10 \log_{10} \frac{\int_{-\infty}^{+\infty} r'^2(\tau) W^2(t-\tau) d\tau}{\int_{-\infty}^{+\infty} l'^2(\tau) W^2(t-\tau) d\tau} \quad (3)$$

$$W(s) = -T_p(1-\omega)\exp(-s/T_p) - \omega T_s \exp(-s/T_s) \quad (4)$$

T_p, T_s は定数で $s \leq 0$ と $s > 0$ で値が異なり, $s \leq 0$ でそれぞれ 3.3 ms, 18 ms, $s > 0$ でそれぞれ 1.8 ms, 9.6 ms, また, ω は -37.25 dB とした。

ここで, 反対側の耳からの抑制を考慮するために,

$$F(t) = \begin{cases} 1 - \exp[-(t-t_0)/T_{inh}] & t_0 \leq t \\ 1 & t_0 \geq t \end{cases} \quad (5)$$

の重み関数を乗じたものを用いる。ここで t_0 は反対側に直接音が到達した時刻とする。すなわち, 右耳の信号の場合であれば

$$r'(t) = r(t) \cdot F(t) \quad (6)$$

となる。また, T_{inh} は抑制の時定数を表し, 過去の心理実験データに基づき 10 ms とした。

また, 両耳に与えられた音のエネルギーを表現するため,

$$BES(t) = 10 \log_{10} \int_{-\infty}^{+\infty} [r^2(\tau) + l^2(\tau)] W^2(t-\tau) d\tau \quad (7)$$

を算出した。なぜなら, 先の ILD は左右耳信号の平均レベルには依存しない表現となるが, 仮に ILD 値が同じでも左右両耳全体のレベルが大きい場合とほとんど聴き取れないレベルでは, 聴感

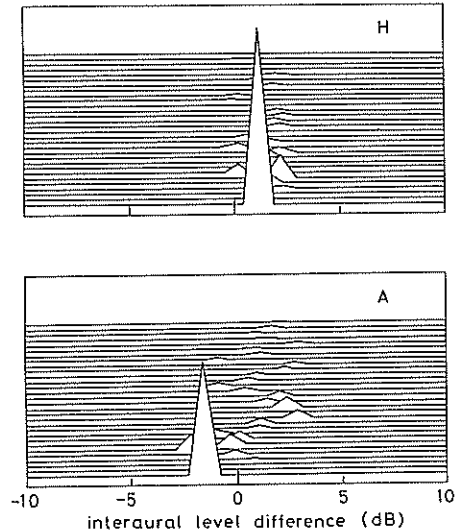


図 9. LSO における機能を表すための Short term interaural level difference (ILD) の例 (ホール H と A). 中心周波数は 4 kHz で, 直接波から 150 ms まで 5 ms ごとに求めた。

上の効果が異なると考えられるためである。

図 9 はこれらを表したものであるが, ピークの高さは BES 値を表し, 横軸に ILD 値を示してある。ここでは, 中心周波数 4 kHz のデータを 5 ms ごとに結果をプロットしてあるが, ホール A と H を比較してみると, 直接波のエネルギー値や ILD 値の時間変動のパターンに大きな違いが見られる。特に, H では直接波のエネルギーが強く反射波部分の ILD 変化がほとんどないのに対し, A では直接波より遅れてやや強い反射波エネルギーが到来しているとともに, その近傍での ILD が大きく変動していることが分かる。

これらのデータ表現と主観評価実験結果との定量的関係を見いだすため, 次のような諸物理量を算出した。

(1) ILD 変動の標準偏差 SD_L

$$SD_L = \left[\frac{\sum_n^k \{ILD(n) - \overline{ILD}\}^2 \cdot BES(n)}{\sum_n^k BES(n)} \right]^{1/2} \quad (8)$$

ここで \overline{ILD} は ILD の平均値を表す。

(2) RIACC ピーク位置の時間変動の標準偏差 SD_T

$$SD_{\tau} = \left[\frac{\sum_n^k (m_p - m_p)^2 \cdot \psi(n, m_p)}{\sum_n^k \psi(n, m_p)} \right]^{1/2} \quad (9)$$

ここで n は離散的時刻を表し、 m_p は時刻 n における RIACC のピーク値に対応した相関遅延時間、 m_p はその平均値とする。

(3) $BES(n)$ の時間平均 BES_M

$$BES_M = \frac{\sum_i^k BES(i)}{k} \quad (10)$$

これらの諸量を 63 Hz から 8 kHz までの 1 オクターブごとの中心周波数帯域で求め、重回帰分析を行った。

まず、主観評価の分析結果で、広がり感などの空間性を表す第一因子と SDL, SDT との重相関係数は 0.84 で比較的高い相関を示した。また、重回帰係数を調べると 1 kHz の SDT, 4 kHz 以上の SDL が正の相関を示し、これらが主観的空間性を増加させる要因であることが分かった。一方、250 Hz 以下の SDT, SDL は共に第一因子と負の相関があった。

従来の聴覚理論では、1.5 kHz 以下では ITD が、1.5 kHz 以上では ILD が音源検知のための手がかりといわれている。特に、ITD の弁別しきい値は 1 kHz 付近で最も小さい値をとり、この帯域で ITD 検出能力が最も卓越していることが知られている。これらの事実を併せて考えれば、1.5 kHz 以上の高域で ILD の変動が広がり感を増加させるのに対し、1.5 kHz 以下では特に 1 kHz 近傍の ITD 変動が重要であることは納得できよう。

また、第二因子に相関が高かったのが、250 Hz ~ 1 kHz の BES_M で、その重相関係数は 0.7 であった。第二因子はおもに音色に関する表現語が多いが、今回相関の高かったこれらの帯域は音楽信号のパワーが最も大きい帯域であり、音色に重要な影響をもたらすためと考えられる。

以上のように、HATS で計測したホールの両耳インパルス応答に、聴覚神経系内での処理過程にならった信号処理を加えた物理量から、主観評価実験の結果を予測できることが分かった。残った因子についても同様の手順により検討を加えた

が、ここではホール音場評価の際に多くの受聴者が重要とみなしている音場空間性と、これに密接な関連を持つ両耳間差の時間変動との関係を中心に述べた。

まとめ

聴覚神経系内の処理に基づく両耳モデルを構成した。このモデルでは、上オリーブ核における LSO, MSO を模擬する機能が加えられている。このモデルの有効性を検討し、各種音場評価への応用の手始めとしてコンサートホール音場評価への応用を試みた。その結果、本モデルを用いれば両耳信号から、音場評価に重要な主観的空間性などがある程度予測できた。特に、音場の主観的空間性を増す物理要因としては、高域では 4 kHz 付近の ILD 変動、1.5 kHz 以下の低域では 1 kHz 付近の ITD 変動が重要であることが分かった。ただし、これらの結果は音楽の種類にも依存して変化すると考えられるため、音楽ソースを変えて検討する必要がある。

また今後は、上オリーブ核以上の神経ネットワークにおける信号処理のモデル化をおこない予測精度を高めることと、各種音場への応用、例えば録音スタジオの音場評価、自動車や航空機騒音評価などの応用を検討したい。さらに、各種異常音の検出など、従来人間が聴感をたよりに行ってきた作業の自動化などへも応用できればと考えている。

本研究を遂行するに当たって多大なご教示、ご助力を賜った黒田勝也氏（キャニオン・レコード）、穴澤健明氏（日本コロムビア）、若林俊介教授（尚美学園短期大学）、利光平大氏（松下通信工業）、山口公典氏（小野測器）、秋山太一郎氏、白井哲也氏（高研）に厚く感謝する。

また、ご助言をいただいた本学浜田晴夫助教授をはじめ研究室関係各位に深謝する。とりわけ、音場計測や主観評価実験で多大な労力を提供してくださった平成元年度本学卒業研究生の増喜太郎君、飯島みのりさんに心より感謝するしだいである。

発表論文

1) 岡部馨, 水内崇行, 浜田晴夫, 三浦種敏: バイ

- ノール計測システム, 電子情報通信学会技術報告, EA89-35, 25-32 (1989).
- 2) Okabe, K., Mizuuchi, T., Hamada, H. and Miura, T.: Monitor-room acoustics and its analysis using head and torso simulator, Proc. 13th ICA (Belgrade), pp. 203-206 (1989).
 - 3) Okabe, K., Mizuuchi, T., Hamada, H. and Miura, T.: Dummy head technique supported by DSP and its application to room acoustic research, Proc. IEEE ASSP Workshop (New York) (1989).
 - 4) 水内崇行, 岡部馨, 浜田晴夫, 三浦種敏: OSS を用いたホール音場評価の検討 その1 —主観計測の結果について—, 日本音響学会研究発表会資料, 1-7-3 (1990. 3).
 - 5) 水内崇行, 岡部馨, 浜田晴夫, 三浦種敏: OSS を用いたホール音場評価の検討 その2 —主観評価と物理特性との対応—, 日本音響学会研究発表会資料, 1-7-4 (1990. 3).
 - 6) 岡部 馨: ダミーヘッドを用いた音場再生, 日本音響学会誌, 46, 650-656 (1990).
 - 7) 岡部 馨, 三浦種敏: 基準的音響伝送系 OSS のための HATS 構成, 日本音響学会誌 (1990), 掲載予定.