聴覚フィルタを用いた室内音場の特徴抽出に関する研究 *

島田　沙織 (九州大)　　鈴木　久晴 (九州大)　　尾本　章 (九州大)

1　はじめに

小さな閉空間の音場の評価を考える際に, 大きな空間で用いる指標をそのまま用いることは難しい。小さな空間は反射波の影響が大きく, 大きな空間の音場よりも複雑であるため, 大空間で用いる指標よりも, より適応範囲の広いモデルを考えることが有用であるだろう。ここでは, 音場の類似度に着目し, その特徴で分類を試み, これを足がかりとして有効な指標を提案することを目標とする。より精度の高い分類法を考えるには，室の特徴を私たちの主観に沿った方法で処理する必要があると思われる。そこで今回は, 室内音響を考える際に, 聴覚モデルを組み込むことがどれくらい有効であるか, 基礎的な検討を行った。

2　音場評価の手法について

室内の音場の特徴を知るために, 従来さまざまな指標が提案されている。これは, 音場を予測するために用いられ, 室内音響設計において何を目標とすればよいかという目標設定にも用いられている。室内音響設計では, 1) 邪魔な騒音がないこと, 2) 言葉が明瞭に聞き取れること, 3) 音楽が美しく豊かに響くこと, 4) 室全体で音場の分布がよいこと, 5) 特異現象がないこと, といった目標値を物理指標によって与えられている。1) に関してはさまざまな方法が提案され, 施行されており, 4)・5) に関しては, 室の形状の工夫や材料の配置を換えることである程度対処することが可能である。しかし, 2)・3) に関しては, とくに人間の主観が入る領域であり, 設計・評価が非常に難しい。また, 一般的に知られている指標には, ストレングス・初期減衰時間・C・D 値などがあるが, それらはインパルス応答をもとに演算で算出できるものである。これらの物理指標はホールのような大きな空間でよく使われており, 必ずしも主観に完全に対応したものであるとは限らない。また, 指標の中に, 直接的に聴覚モデルを組み込んだものはほとんど無く, 物理指標と聴覚モデルを結びつける研究も積極的に進められているとは言い難い。

一方, 主観との対応を考えるにあたって, 聴覚モデルが研究されてきた [1]。主観との対応を考えるには, 私たちがどのようにして音を聞き取っているかを考える必要があるだろう。音が耳から入ってきて神経に到達するまでの過程, これを低次な器官の特徴と呼び, 脳が時間・周波数に依って違う捕らえ方をするだろうという考え方を, 高次な器官の特徴と呼ぶと, 聴覚モデルは, 聞こえが低次な器官に依存すると考え分析を行うものである。

聴覚モデルは, Zwicker の知覚ラウドネスモデルにより表されており, 耳の特性を組み込んだ量である心理量で処理を行う。このモデルを用いた方法では, 耳の特性である, 基底膜の特性のなかでも時間マスキング・周波数マスキングといったものを考慮することができる。これまでに，この聴覚モデルは，ノイズリダクション [2] や mp3 などの情報圧縮などの研究で有効であることが示され，よく用いられてきた。特に周波数マスキングに多くの影響を受ける基底膜の特性を組み込むことで，より精度高いノイズ除去を行うことができ，それはつまり，ノイズリダクションにおいては，聴覚モデルを用いた処理が，より主観に対応していることを示している。ということは，ノイズリダクションや情報圧縮においては，人の音の聞き取りは，低次な器官に依存していたということになる。このような分野においては, 低次な器官の特徴を考慮する聴覚モデルの有用性が認められてきたが, 音場の評価においては, その研究はあまり進

---

* Extractiong Features of the Sound Fiels in Rooms by using Auditory Filter.by Saori SHIMADA, Hisaharu SUZUKI, Akira OMOTO(Kyushu University)

んでいない。

小さな閉空間の分類を行う際に聴覚モデルを適用し, これが, よりよい主観との対応をもたらすものであるならば, より精度の高い小空間音場の分類も可能になるのではないかと考えられる。そこで, 今回は, 聴覚モデルの音場への適用を考える前に, まず, モデルがどれくらい主観に対応しているのかということを, 主観評価実験との結果と比較することで調べることにする。

3 心理音響モデル

物理量である時間領域での音圧を心理量に変換するための手順は以下のようになる [3]。離散信号 $x(n), 0 \leq n \leq N-1$ として, n は時間領域のインデックスとする。これを K サンプル含むフレーム長で分解し, 周波数領域に変換する。その際, オーバーラップを考慮してもよい。$x(n)$ をフレーム長 K, オーバーラップ長 L で, 短時間フーリエ変換 (STFT) したものを $X_w(k,i)$ とする。

$$X_w(k,i) = \sum_{n=0}^{K-1} x(n + off_i) w(n) I_K^{kn} \qquad (1)$$

$0 \leq k \leq K-1$　　$I_N = \exp(-j\frac{2\delta}{N}n)$

K：フーリエ変換の長さ

$w_i(k)$：ハニング窓　　$off_i = (K-L) \times i$

i：窓の index　　k：周波数領域の index

次にパワースペクトル $X_p(k,i)$ を算出する。

$$X_p(k,i) = \mid X_w(k,i) \mid^2 \qquad (2)$$

この $X_p(k,i)$ により, 臨界帯域に対応するバーク領域でのパワーを求める。

$$X_a(b,i) = a_0(b) \sum_{k=k_{lb}}^{k_{hb}} X_p(k,i) \qquad (3)$$

$0 \leq b \leq B-1$　　$B = 25$

k_{lb}：周波数帯域の下限

k_{hb}：周波数帯域の上限

$a_0(b)$：Outer-inner ear transfer function $a_0(b)$ は, 周波数帯域間の間隔と各帯域ごとの周波数成分の数に依存する関数である。

続いて, $X_a(b,i)$ から, 時間マスキングを考慮した量 $X_t(b,i)$ を算出する。時間マスキングでは, すべての過去のフレームの影響を受ける。

$$\begin{aligned} X_t(b,i) &= X_a(b,i) + T_f(b) X_t(b.i-1) \\ &= X_a(b,i) + X_t(b,i-1) e^{T_f(b)} \end{aligned} \qquad (4)$$

$T_f(b) = -\frac{d}{\tau(b)}$

d：隣接した短時間フレーム間の距離

$\tau(b)$：マスキング実験から得られた関数

さらに, 周波数マスキングについて考慮した値 $X_f(b,i)$ を求める。周波数マスキングでは, 臨界帯域上で両隣から影響を受ける。

$$X_f(b,i) = \Big[\Big\{ \sum_{v=b}^{B-1} [S_1(v-b) X_t(v,i)]^{\frac{\delta}{2}} \Big\} \qquad (5)$$

$$+ \Big\{ \sum_{v=0}^{b-1} [S_2(v,b-v) [X_t(v,i)]^{1+0.02(b-v)dz} \Big\}^{\frac{\delta}{2}} \Big]^{\frac{2}{\delta}}$$

S_1：b より上方の周波数のための関数

S_2：b より下方の周波数のための関数

S_1, S_2：蝸牛の特性関数

ここで求めた $X_f(b,i)$ より, Zwicker と Fastle によるラウドネス関数の表現に従った関数を求める。これは, $X_p(k,i)$ の心理音響量の概念と一致する。

$$X_l(b,i) = \kappa \Big[\frac{E_0(b)}{s} \Big]^{\gamma} \Big\{ \Big[1 - s + s \frac{X_f(b,i)}{E_0(b)} \Big]^{\gamma} - 1 \Big\} \qquad (6)$$

κ：任意のスケール定数

s：実験で得られるパラメータ

γ：最適化されたパラメータ

E_0：$a_0(b)$ を掛け合わせた各周波数帯ごとの絶対最小可聴値のエネルギー

4 モデルの適用－分析方法－

心理音響モデルを適用し信号の類似度を算出する過程は, 以下のようなものである。

1. 二つの信号を, 物理量 X_ω から心理量 X_l に変換・算出する。
2. バーク領域で, 二つの心理量間の距離を求める。
3. 各信号間の距離を算出し, 一次元マップを得る。

信号の類似度の算出には, 以下のような距離尺度を採用し, どの距離尺度を用いて分析するのがより主観に対応するかを検討した。これらは, パターン認識によく用いられる距離尺度である。

- ユークリッド距離
- マハラノビス距離
- 内積を用いた類似度

ユークリッド距離

差の二乗和の平行根で, いわゆる n 次元の「直線距離」である。連続距離としてはユークリッド距離が広く用いられる。

$$d^2(f,g) = \| f-g \|^2 = \sum_{m=1}^{B}(f_m - g_m)^2 \quad (7)$$

マハラノビス距離

各変量の主成分得点を分散 1 に基準化したユークリッド距離。

$$d^2(f,g) = (f-g)^T C^{-1}(f-g) \quad (8)$$

C：共分散行列

内積を用いた類似度

類似度は, 距離が異なり, 2 つのベクトルの内積を評価し, 2 つのベクトルのなす角度が小さいもの, つまり類似度が 1 に最も近いものがもっとも似ているとする。

$$S(f,g) = \frac{(f,g)}{\| f \| \| g \|} = \frac{\sum_{m=1}^{B} f_m g_m}{\sqrt{\sum_{m=1}^{B} f_m^2}\sqrt{\sum_{m=1}^{B} g_m^2}} \quad (9)$$

これらの尺度を用いて二つの心理量間の類似度である距離を求めるのであるが, このときフレーム分けによる窓の数を I とすると, 二つの心理量 X_l は, $I \times B$ の行列となっている。距離は, フレーム毎に求める。つまり, 一組につき I 個の距離が出るのである。ここでそれぞれの距離に対して, I 個のデータの平均値と最大値, 内積を用いた類似度に対しては最小値をとり, 各信号間の距離を算出し, 一次元マップで表した。

5 主観実験の方法

一対比較法で類似度判断を行った。二つ一組で曲を聴き, 「似ている」か「似ていない」かを 7 段階の尺度で判断した。刺激にはドライソースに, オーディオを聞くための部屋のインパルスレスポンスを畳み込み, さらにイコライザを用いて七つの刺激を作製した。イコライザに関しては, まったくかけてないもの, 高域 (2k-8kHz)・中域 (250-1kHz)・低域 (63-125Hz) でそれぞれ 8dB ずつ音量を上げたものと下げたものの七種類である。被験者は日常生活に支障のない聴力を有する, 九州大学の 22 歳〜26 歳までの学生（男性 8 名, 女性 4 名）である。

得られたデータにより, 一次元でその類似度のマップを得た。

6 モデルを用いた分析と主観実験結果との比較

Fig.1 は, 主観評価実験の結果である。これは, 刺激間の類似度を得点として, 各刺激間の距離を一次元のマップ上にあらわしものである。図中には, 刺激にかけたイコライザの特性を, たとえば低い周波数帯でレベルを落としたものを low down, 元の素材を sorce のように示している。

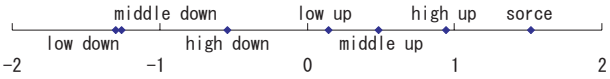

Fig.1 Result of the subjective test

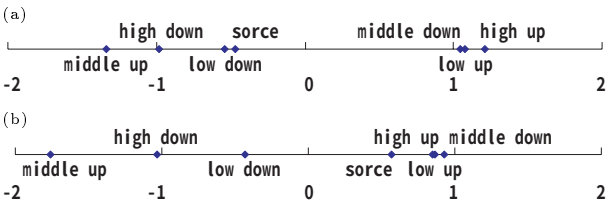

Fig.2 Arrangement of each stimulus by Euclid distance. (a)Average (b)Maximum

Fig.2 は, ユークリッド距離から出した結果である。Fig.2 の (a) は, 各フレームごとに距離を算出して, 平均をとったものである。Fig.2 の (b) も, ユークリッド距離から出した結果であるが, こちらは, 各フレームごとに距離を算出して, その中で最大値をとったものである。Fig.2 では, 平均値も最大値も同じ方法で距離を算出したことにより, やや類似しているのだろうと考えられる。

Fig.3 の (a) は, ユークリッド距離を応用した算出方法であるマハラノビス距離から出した結果である。これは, 各フレームごとに距離を算出して, 平均をとったものである。

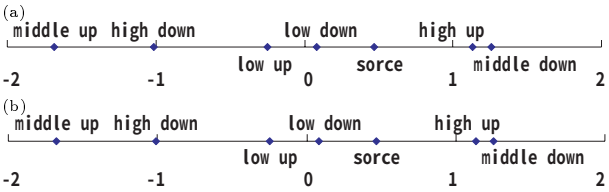

Fig.3 Arrangement of each stimulus by Mahalanobis distance. (a)Average (b)Maximum

Fig.3 の (b) も, マハラノビス距離から出した結果であるが, これも, 各フレームごとに距離を算出して, その中で最大値をとったものである。Fig.3 では, 平均値も最大値も同じ方法で距離を算出したことにより, 類似している。さらに, 分散を考慮した方法であるので, 平均であるか, 最大値であるかの値の処理の方法によらず安定した結果を得られているのではないかと思われる。

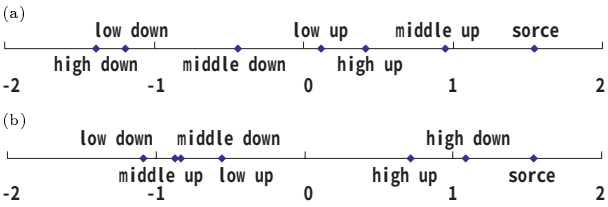

Fig.4 Arrangement of each stimulus by inner product. (a)Average (b)Maximum

Fig.4 の (a) は, 内積を用いた類似度により算出した結果である。これは, 各フレームごとに距離を算出して, 平均をとったものである。Fig.4 の (b) も, 内積を用いた類似度から出した結果であるが, 各フレームごとに距離を算出して, その中で最大値をとったものである。Fig.4 の平均値と最大値は同じ方法で距離を算出しているが, よく類似しているとは言い難い。

また全体的に見て, 内積を用いた類似度の平均での結果は主観評価実験の結果と, やや対応がとれているように見える。今回の結果より, 内積を用いた類似度の平均での結果を用いると, 主観に比較的対応した結果が得られるのではないかと考えられる。

7 考察

主観評価実験の結果では, 中域・低域の周波数帯のレベルを落としたもの各々が特に元の刺激と似ていないというのが結果よりわかる。また, 高域の周波数帯のレベルを上げたものが, 元の刺激に似ている, つまりレベルの操作があまり影響していないのではないかと考えられる。

これと比較して, 内積を用いた類似度では, 平均値と最大値で比較的違いが出たが, 平均値の方で, 高域・中域・低域でレベルを上げたものそれぞれが元の刺激との類似度が高く, 高域・中域・低域でレベルを下げたものそれぞれが類似度が低いという点で, 主観評価実験の結果に比較的類似した結果が得られた。またこの方法で得られた二つの結果を平均することにより, さらに主観評価実験の結果に似てくることがわかった。それが Fig.5 である。

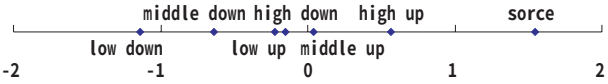

Fig.5 Arrangement of each stimulus by inner product.

つまり, 内積を用いた類似度は, 平均値に最大値が及ぼす影響をある程度の重みとして付加すると, 人間が判断する類似度に近しいものが得られるのではないかと考える。

8 今後の展望

今回の研究によって, 周波数帯でレベルをを変化させた刺激に対しては, 類似度の算出方法しだいでは主観に比較的対応した結果が得られるのではないかと思われる。これからは, 素材の種類を増やして主観評価実験を行うことで精度を高め, 一方で残響時間を変化させた刺激でも検討を行う必要がある。

参考文献

[1] John G.Beerends,*et al.*: A Perceptual Audio Quality Measured Based on a Psychoacoustic sound Representation, J.AES.,40(12), pp963-978(1992)

[2] Dionysis E.Tsoukalas,*et al.*: Perceptual Filter For Audio Signal Enhancement, J.AES.,45(1/2), pp22-36(1997)

[3] James D.Johnston,*et al.*: Transform Cording of Audio Signals Using Perceptual Noise Criteria, IEEE Journal on selected AREA in Communications.,6(2), pp314-323(1988)