スピーカ間の音量差に基づく音像生成手法における個人適応の検討*

☆天野成祥，山田武志，牧野昭二，北脇信彦（筑波大）

1　はじめに

近年のタッチパネルディスプレイの普及は目覚しく，身の周りの様々な場面で目にするようになっている．しかし，視覚に障がいを有する人にとっては，タッチパネルディスプレイの操作が本質的に難しいという問題が生じている．我々はこの問題を解決するために，音像生成手法を用いたタッチパネルディスプレイを提案している[1]．音像を画面上に生成し，画面に表示されている内容を空間的な音情報として提示することにより，ユーザは視覚に頼ることなく聴覚のみによってタッチすべき場所を把握することが可能となる．

一般に，音源から両耳までの音響伝達特性を厳密に再現するほど定位精度は高くなる．しかしその一方で，ヘッドホンの使用やユーザの頭の位置の固定といった強い制約を課すことになる．このような制約を緩和するために，画面四隅に設置したスピーカ群を用いることとする．

小澤らは，スピーカ群の音量の組み合わせ（以降音量設定）と人間による定位位置の関係を実測し，それに基づいて指定位置に定位させるための音量設定を推定するという手法を提案している[2]．この手法では，人間は水平方向の定位と垂直方向の定位を独立に行うと仮定し，スピーカ群の音量設定と人間による定位位置の関係を水平方向，及び垂直方向において独立に測定している．しかし，それぞれの方向において定位を独立に行うという仮定は必ずしも成り立たないことが知られている．

そこで我々は，水平方向と垂直方向の両方のスピーカ群の音量設定と人間による定位位置を測定し，その関係をモデル化した．そして，そのモデルに基づいて指定位置に定位させるための音量設定を推定する手法を提案した[3]．しかし，音像の定位位置の個人差により，音量設定の推定精度が低下することが分かった．

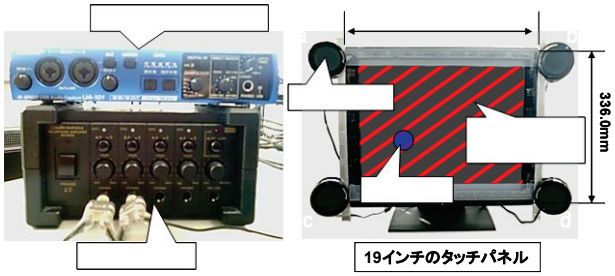

Fig.　1: System overview

本稿では，この問題に対処するために，構築したモデルに対して個人適応を行う手法を提案する．

2　提案手法

2.1 システムの概要

本研究で用いているシステムの概要を Fig. 1 に示す．本システムは，PC，オーディオインタフェース，アンプ，タッチパネル及び画面の四隅に設置された小型スピーカで構成されている．以上の機器と音像定位技術を用い，画面上の指定位置に音像を生成することにより，ユーザにタッチすべき位置を知らせる．ここで，画面の大きさは 448mm×336mm(19 インチ)である．またタッチできる範囲の解像度は 1024pixel×690pixel である．なお，以降の記述において，画面上部のスピーカ a，b から出る音を『上方音源』，画面左部のスピーカ a，c から出る音を『左方音源』などと表現する．本システムでは各スピーカ間の音量差を用いて，音像の位置を制御する．なお，音量の調節は次のような方針により行う．

- 垂直方向の位置の制御：上方音源，下方音源のいずれかの音量を減らす．
- 水平方向の位置の制御：右方音源，左方音源のいずれかの音量を減らす．

2.2 音量設定と定位位置の関係を表すモデル

音量設定と人間による定位位置の関係を被験者実験により測定し，モデル化を行なった[3]．

実験条件を Table 1 を示す. 使用する音源は，

* A study on personal adaptation for a sound image control method based on sound level differences among loudspeakers， by Shigeyoshi Amano， Takeshi Yamada， Shoji Makino and Nobuhiko Kitawaki (University of Tsukuba).

男性による『学生』という音声である．また，音量設定は画面を満遍なくタッチできるような 72 パターンを用いた．

測定結果を Fig.　2 に，その測定結果から構築したモデルの一部を Table 2 に示す．Fig.　2 において，縦軸と横軸は各々垂直方向と水平方向の定位位置をピクセルにより示している．なお，各点は 72 通りの音量設定の各々に対して被験者 10 名が定位した位置の平均値であり，エラーバーは 95%信頼区間である．また Table 2 は，各音量設定に対する定位位置の平均値と 95%信頼区間片側幅を示している．このうち音量設定と定位位置のペアからなるデータの集合をモデルと称している．

Fig.　2 より，おおよそ画面上を網羅するように定位していることが確認できる．画面の端と下方への定位が少ないが，前者については，被験者が画面の枠に触れるのを避けようとしていること，後者については，そもそも下方に定位することが難しい[4]ということが原因と考えられる．

次に，Fig. 3 に定位位置の被験者間のばらつきの例を示す. 図中の被験者 10 名の平均値は，Fig. 2 の丸で囲った 1 点に対応する．被験者 A の平均値は，被験者 B の平均値と比べ，被験者 10 名の平均値から大きく離れていることが見て取れる．このことから，定位位置には個人差があることが確認できる．本手法では構築したモデルを用いて指定位置に定位させるための音量設定を推定することから，定位位置の個人差を解消しない場合，指定位置付近を定位させることが困難になると考えられる．従って本稿では，構築したモデルに対して個人適応を行い，個人差の軽減を目指す．

2.3 個人適応

本稿では，まずシステムを利用する前のキャリブレーションに相当するオフライン個人適応について検討する．まず，不特定モデル(多数の被験者によるモデル)中の数点の定位位置における音量設定を用いて音像を生成し，それぞれの音像に対する利用者の定位位置を適応データとして取得する．その適応データを用いて，上記の数点の定位位置については MAP 推定法，それ以外の全ての定位位置については移動ベクトル場平滑化法（以降，MAP-VFS）を適用することにより，個人適応を行う．

以下において，個人適応の流れを説明する．まず，適応データの取得対象である不特定モデル中の定位位置に対して，次式により MAP 推定による個人適応を行う．

Table 1: Experimental conditions

音源	『学生』という音声(男性)
画面からユーザまでの距離	約 45 cm
音量設定のパターン数	72 パターン
被験者数	10 人
試行回数	各音量設定に対し 5 回
回答方法	定位位置をタッチ

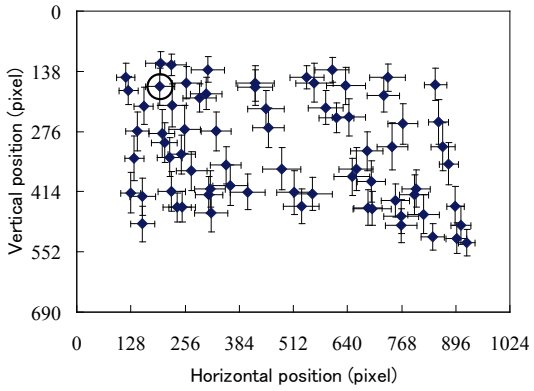

Fig.　2: Positions localized for each volume settings

Table 2: Model of the relationship between localized positions and volume settings

音量設定(dB) 左上，右下 左下，右上	定位位置 (pixel) (水平方向, 垂直方向)	95%信頼区間幅 (pixel) (水平方向，垂直方向)
50, 44 41, 35	(130, 140)	(24, 21)
50, 44 44, 38	(147, 131)	(27, 33)
50, 44 47, 41	(188, 146)	(27, 44)

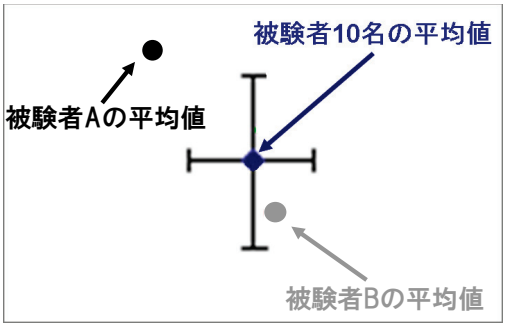

Fig.3: Individual variation for a localized position

$$\hat{\mu}_i = \frac{n}{n+\tau}\boldsymbol{m} + \frac{\tau}{n+\tau}\boldsymbol{\mu}_i \qquad (1)$$

$\hat{\mu}_i$ は個人適応後の定位位置を示している．$\boldsymbol{m}$ は適応データの定位位置の平均値であり，$\boldsymbol{\mu}_i$ はモデル中の定位位置である．また，n は適応デ

ータのサンプル数であり，τはモデル中の定位位置の確からしさを示すパラメータである．τの値が大きい場合，個人適応後の定位位置はモデル中の定位位置に近い位置に設定され，τの値が小さい場合，個人適応後の定位位置は，逆に適応データの定位位置に近い位置に設定される．

次に，適応データの取得対象ではない不特定モデル中の定位位置に対して，MAP-VFS による個人適応を行う．適応データの取得対象である定位位置の移動方向と移動量を示す移動ベクトル$\boldsymbol{v}_i$は次式により求めることができる．

$$\boldsymbol{v}_i = \hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i \tag{2}$$

このとき，適応データの取得対象でない定位位置の移動ベクトルの推定値$\hat{\boldsymbol{v}}_j$は以下の式により求められる．

$$\hat{\boldsymbol{v}}_j = \frac{\sum_{i=1}^{k}\left(\boldsymbol{v}_i \times w_{ij}\right)}{\sum_{i=1}^{k} w_{ij}} \tag{3}$$

kは，適応データの取得対象となる定位位置の個数を示しており，$\boldsymbol{v}_i$は適応データの取得対象となる定位位置に対して個人適応を行なった際に，式(2)によって求められた移動ベクトルである．w_{ij}は，適応データの取得対象ではない定位位置$\boldsymbol{\mu}_j$から$\boldsymbol{\mu}_i$までの距離によって決まる重み係数である．本稿では，重み係数を次式により定めている．

$$w_{ij} = \exp(-d_{ij} / s) \tag{4}$$

d_{ij}は$\boldsymbol{\mu}_i$と$\boldsymbol{\mu}_j$の間のマハラノビス距離である．sは平滑化の度合を制御する係数であるが[5]、本稿では平滑化を行わないため，d_{ij}の影響を調整するパラメータとして用いることとする．

3　提案手法の有効性の検証

3.1 個人モデルと個人適応後モデルの比較

本稿では，10 名中 1 名のデータから構築したモデルを個人モデル，残り 9 名のデータから構築したモデルを不特定モデルと呼ぶこととする．個人モデル及び不特定モデルは，クロスバリデーションにより，10 パターン用意する．

前節において述べた手法を用いて，不特定モデルに対して個人適応を行い，適応後のモデルと個人モデルの類似度（以降，適応後類似度）を測定する．そして，適応前のモデルと個人モデルの類似度（以降，適応前類似度）と比較することにより，個人適応の有効性を検証する．なお，適応前（後）類似度には個人適応前（後）の不特定モデル中の定位位置から個人モデル中の定位位置までの距離の平均を用いる．また，前節の式(1)におけるパラメータτ, 及び式(4) おけるパラメータsについては，事前検証により最も良い結果が得られた$\tau = 2$，$s = 10$を用いることとする．

まず，不特定モデル中の四隅にある定位位置を適応データの取得対象として個人適応を行ない，類似度を求めた．これは，不特定モデルの四隅に位置する定位位置の移動ベクトルを求めることでモデル全体を拡大，縮小，もしくは左右上下に移動させやすいと考えられることによる．

検証結果を Table 3 に示す. 各類似度の値は，10 パターンの個人モデルと不特定モデルの組み合わせにおいて，類似度をそれぞれ求めた際の平均値を示している．適応後類似度の方が適応前類似度よりも値が小さくなっていることが見て取れる．

次に，適応後類似度をさらに向上させるために，不特定モデル中の四隅にある定位位置に加えて，他の数箇所の定位位置を適応データの取得対象として追加する．なお，追加する定位位置については，適応後類似度の値が一番小さくなるものを選ぶ（画面中央の定位位置が選ばれる傾向があった）．

検証結果を Fig. 4 に示す．横軸は適応データとして用いる定位位置の数（四隅を含む），縦軸は適応後類似度を示している．適応データの取得対象である定位位置の数が多くなるほど，適応後類似度の値が小さくなっていることが見て取れる．しかし，適応データの取得対象である定位位置の数が 6 以上の場合には改善度合が低いということもわかった．従って本稿では，個人モデル中の四隅を含む 5 箇所の定位位置を適応データとして用いることとする．

3.2 音量設定の推定における有効性

個人適応前，個人適応後の不特定モデルを用いて，指定位置に対する音量設定を推定する．本稿では推定誤差によって有効性を検証するため，音量設定の値が既知である個人モデル中の定位位置を指定位置として用いることとする．音量設定の推定値については，個人適応前，個人適応後の不特定モデルにおける音量設定の重み付け和（指定位置から近い

Table 3: Comparison of the similarity

適応前類似度（平均）	適応後類似度（平均）
177.6 pixel	130.27 pixel

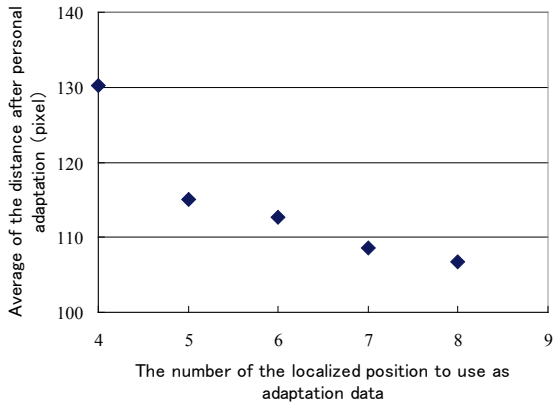

Fig.4: Relationship between the similarity and the number of the adaptation data

ほど重みを大きくする）により決定する．

10 パターンの個人モデルにおける全定位位置の各々を指定位置とした場合の検証結果を Table 4 に示す．各推定誤差の値は，10 パターンの各誤差の平均値を示している．個人適応後の不特定モデルを用いた場合の方が，個人適応前の不特定モデルを用いた場合よりも推定誤差が小さいことが見て取れる．従って，個人適応後の不特定モデルは音量設定の推定に対して有効であると考えられる．

しかし，垂直方向における推定誤差は水平方向における推定誤差と比べて値が大きいという問題がある．その原因の 1 つとして，モデル中の外れ値の存在が挙げられる．例えば，音量設定 A よりも音量設定 B の上方音源の音量が大きい場合，音量設定 A を用いた場合の方が上方の位置を定位すると考えられる．しかし，Fig. 5 のように逆に下方向を定位してしまっているという場合が存在している．そこで，このような外れ値を除外したモデルを用いて音量設定を推定することにより，推定精度の向上を試みた．

検証結果を Table 5 に示す．Table 4 に示した結果と比べて，垂直方向の音量設定推定誤差が小さくなっていることが見て取れる．従って，今後は外れ値を除外した不特定モデルを用いて個人適応，及び音量設定の推定を行う必要があると考えられる．

Table 4: Estimated error of the volume setting

	水平方向音量設定推定誤差 (dB)	垂直方向音量設定推定誤差 (dB)
個人適応前モデル	1.25	4.39
個人適応後モデル	0.93	3.79

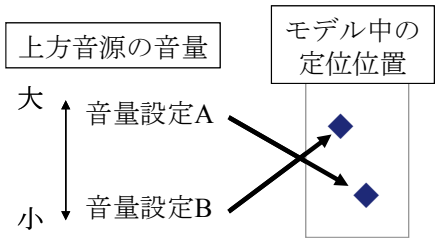

Fig. 5: Example of the outlier

Table 5: Estimated error of the volume setting after removing the outlier

	水平方向音量設定推定誤差 (dB)	垂直方向音量設定推定誤差 (dB)
個人適応前モデル	1.24	3.66
個人適応後モデル	0.93	3.12

4 まとめ

本稿では，定位位置の個人差を軽減するために，音量設定と人間の定位位置の関係を表すモデルに対して個人適応する手法を提案し，その有効性を示した．今後は被験者実験を行い，3. 2 節で示した推定誤差が定位精度にどの程度影響があるか検証する予定である．

参考文献

[1] 木村知浩， 山田武志， 北脇信彦， ″タッチパネルのための音像定位インタフェースの検討，″ 電子情報通信学会総合全国大会， pp.338，Mar. 2009.

[2] 小澤賢司， 降矢龍浩， ″合成音像による位置表示が GUI におけるオブジェクト探索に及ぼす効果，″ 情報処理学会論文誌 Vol.42，No.6，pp.1299-1310，Jun. 2001.

[3] 天野成祥， 山田武志， 牧野昭二， ″視覚障がい者のタッチパネル操作支援のための音像生成手法の検討，″ 日本音響学会春季研究発表会，pp.901-902, Mar. 2011.

[4] 降矢龍浩， 小澤賢司，鈴木陽一 "音圧レベル差を制御した合成音像による 2 次元音像定位，" 日本音響学会研究発表会講演論文集，pp.495-496，Oct. 2001.

[5] 益子貴史，田村正純，徳田恵一，小林隆夫，″HMM に基づく音声合成システムにおける MAP-VFS を用いた声質変換，″ 日本音響学会春季研究発表会，pp.2509-2561, Dec. 2000.