

音量差に基づく音像生成における個人適応手法の有効性検証*

☆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きている。しかし、視覚に障がいや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、ディスプレイの4隅に設置したスピーカ群の音量設定と人間による定位位置の関係をモデル化し、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[1, 2].

しかし、音像の定位位置には個人差があり、それにより音量設定の推定精度が低下してしまうという問題がある。我々はこの問題を解決するために、MAP 推定法及び移動ベクトル場平滑化法を用いて、モデルに対して個人適応を行う手法を提案している[3].

本稿では、上記の提案手法により個人適応を行い、適応後のモデルを用いてスピーカ群の音量設定を推定する。そして被験者実験により定位精度に対する提案手法の有効性を検証する。

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4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、画面の大きさは横 438mm×縦 372mm (19 インチ)である。またタッチできる範囲の解像度は横 1024pixel×縦 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 を『上方音源』、画面左部のスピーカ a, c 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。

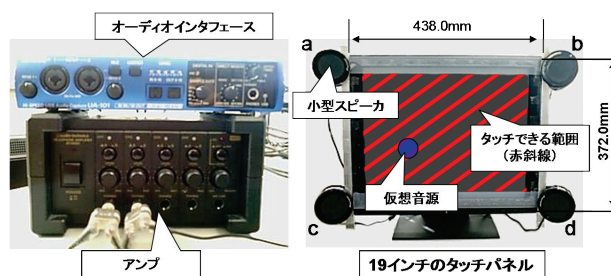


Fig.1: System overview

- 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のモデル化

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[1].

実験条件を Table 1 に示す。使用する音源は、男性による単語音声である。画面からユーザまでの距離は約 45cm であり、音量設定は画面を満遍なくタッチできるように 72 パターンを用いた。また、被験者は 10 名であり、各音量設定に対して 5 回ずつ定位位置をタッチする。なお、音像の提示に用いる音量設定の順番はランダムとした。

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 1 つに対して被験者 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を示している。このような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。

次に、Fig.3 に定位位置の被験者間のばらつきを例を示す。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、定位位置には顕著な個人差があることが確認できる。提案手法では構築したモデルを

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

Table 1: Experimental conditions

音源	単語音声 (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
試行回数	各音量設定に対し 5 回
被験者数	10 人
回答方法	定位位置をタッチ

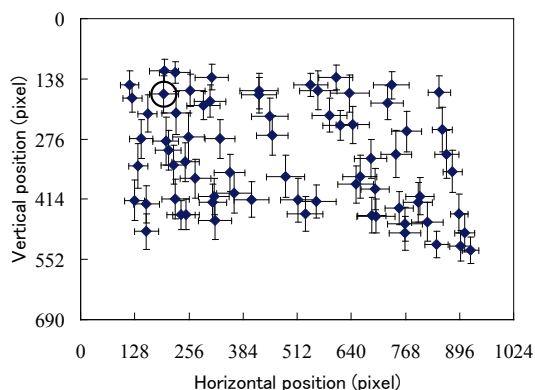


Fig.2: Positions localized for each volume setting

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

音量設定(dB) 左上, 右下 左下, 右下	定位位置(pixel) 水平方向, 垂直方向
50, 44 41, 35	130, 140
50, 44 44, 38	147, 131
50, 44 47, 41	188, 146

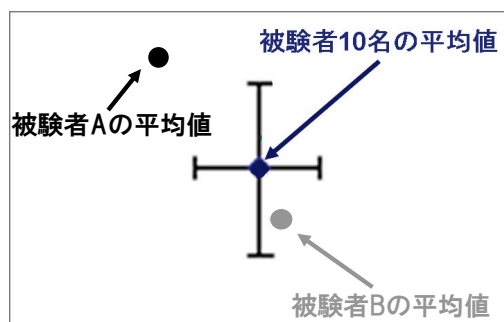


Fig.3: Individual variation for the localized position

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考慮しないと、指定位置に定位させることが困難になると考えられる。そこで、構築したモデルに対して個人適応を行い、個人差の軽減を図る。

2.3 不特定モデルに対する個人適応

我々は、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手法を提

案している[3]. まず、不特定モデル(2.2において構築したモデル)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 (Maximum A Posteriori) 推定法、それ以外の全ての定位位置については MAP-VFS (VFS: Vector Field Smoothing) 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。

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

$\hat{\mu}_i$ は個人適応後の定位位置を示している。 m は適応データにおける定位位置の平均値であり、 μ_i は不特定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 は不特定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。個人適応後の定位位置は、 τ の値が大きい場合は不特定モデル中の定位位置に近い位置に設定され、 τ の値が小さい場合は逆に適応データにおける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。なお、本稿では Fig.4 に示すように、不特定モデル中の適応データの取得対象でない定位位置を μ_j とした場合を例に挙げ、説明する。不特定モデル中の適応データの取得対象である定位位置 μ_i (Fig.4 中では μ_1, μ_2 に相当) の移動方向と移動量を示す移動ベクトル v_i (Fig.4 中では v_1, v_2 に相当) は次式により求めることができる。

$$v_i = \hat{\mu}_i - \mu_i \quad (2)$$

このとき、 μ_j の移動ベクトルの推定値 \hat{v}_j は以下の式により求められる。

$$\hat{v}_j = \frac{\sum_{i=1}^k (v_i \times w_{ij})}{\sum_{i=1}^k w_{ij}} \quad (3)$$

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、 v_i は、式(2)によって求められた移動ベクトルである。 w_{ij} は、 μ_i から μ_j までの距離によって決まる重み係数である。ここで、重み係数は次式により決定している。

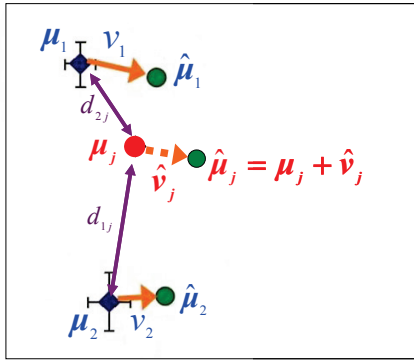


Fig.4: Personal adaptation

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

d_{ij} (Fig.4 中では d_{1j} , d_{2j} に相当) は μ_i と μ_j の間の距離であり, s は平滑化の度を制御する係数である[4].

以上によって推定した移動ベクトル \hat{v}_j を μ_j に足し合わせるにより, 個人適応後の定位位置 $\hat{\mu}_j$ を取得することができる.

2.4 指定位置における音量設定の推定

提案手法では, モデル中の各音量設定の重み付け和によって, 指定位置における音量設定を決定する. 例えば, Fig.5 に示すように丸い点を指定位置とした場合, 指定位置の音量設定 \hat{V} は以下の式によって推定される.

$$\hat{V} = \sum_{i=1}^{72} (V_i \times w_i) \quad (5)$$

V_i はモデル中の各定位位置に対する音量設定 (Fig.5 中では V_1, V_2, V_3 に相当) を示している. w_i はモデル中の各定位位置 μ_i (全 72 パターン) から指定位置 μ までの距離 d_i (Fig.5 中では d_1, d_2, d_3 に相当) に応じた重みを示しており, 以下の式によって決定される. なお, 距離が近いほど重みは大きくなる.

$$w_i = \frac{\exp(-d_i)}{\sum_{i=1}^{72} \{\exp(-d_i)\}} \quad (6)$$

3 提案手法の有効性検証

3.1 実験方法

本稿では, 提案手法を用いて不特定モデルに対して個人適応を行う. そして, 不特定モデル, 個人適応後の不特定モデル (適応モデル), 及び個人による音量設定と定位位置の関係のモデル (個人モデル) を用いて指定位置に対する音量設定をそれぞれ推定する. それら推定した音量設定を用いて被験者実験を行い, 定位精度を比較することによって提案手法の有効性を検証する.

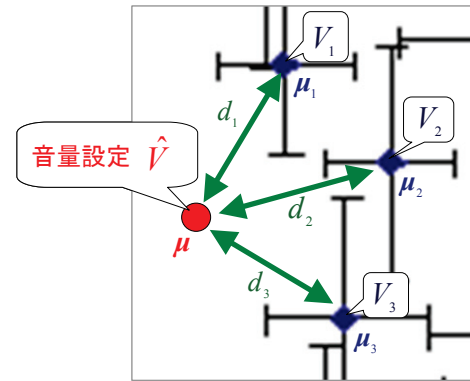


Fig.5: Estimation of a volume setting

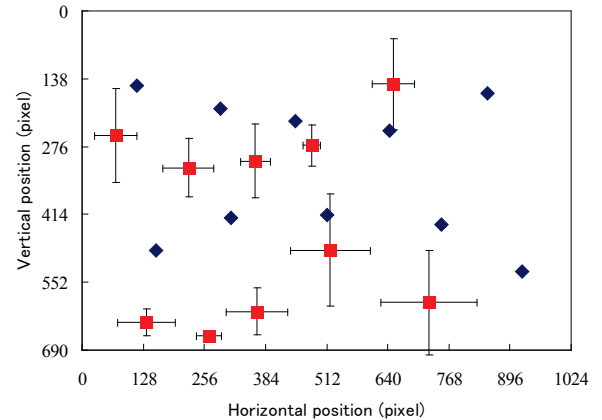


Fig.6: Localized positions for 10 desired positions

まず, 不特定モデル構築の際とは異なる 10 名を被験者とし, 10 点の指定位置に対する定位位置を測定した. その測定結果より, 指定位置から定位位置までの距離の平均値が特に大きい 3 名を個人適応の必要性が大きいと判断した. 例として 3 名中 1 名の定位位置の測定結果を Fig.6 に示す. 菱形の点は指定位置, 四角形の点は, その被験者による定位位置の平均値を示している. 定位位置が指定位置から大きく離れてしまっていることが見て取れることから, 個人適応の必要性が大きいことがわかる.

次に, 不特定モデル中の 5 点 (4 隅の点 + 中央付近の 1 点) の定位位置における音量設定を用いて, 音像を生成する. そして, それぞれの音像に対する各被験者の定位位置を取得する. それを適応データとして用い, 不特定モデルに対して個人適応を行う.

不特定モデル, 個人適応により取得した適応モデル, 及び各被験者による個人モデルを用いて, Fig.7 に示す 12 点の指定位置の音量設定をそれぞれ推定する. その後, 推定した音量設定を用いて, Table 1 と同様の条件下 (但し, 被験者は適応データ取得の際と同じ 3 名) で被験者実験を行う.

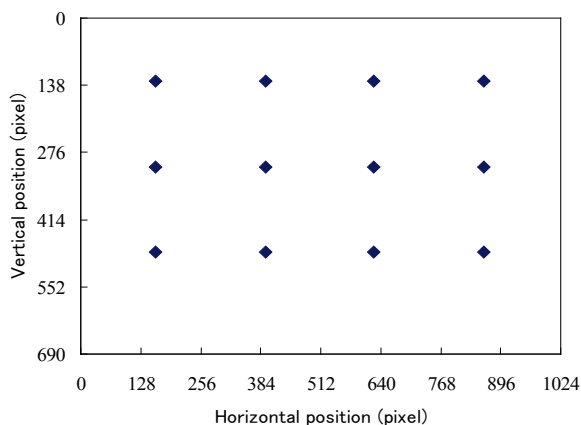


Fig. 7: Twelve desired positions

3.2 実験結果

まず、不特定モデルを用いた場合と適応モデルを用いた場合の実験結果（3名中1名）を Fig.8, Fig.9 にそれぞれ示す。菱形の点は指定位置を、四角形の点は、その被験者による定位位置の平均値を示している。Fig.8 より、不特定モデルを用いた場合は、定位位置が全体的に右上の方向に寄っており、特に画面左側の指定位置に対する定位精度が低くなっていることが見て取れる。一方、適応モデルを用いた場合には、Fig.9 に示すように画面左側の指定位置に対する定位精度が改善できていることが見て取れる。

次に、指定位置と定位位置の距離の平均値を、被験者別に Table 3 に示す。なお、音量設定の推定に用いたモデル別に値を示しており、値が小さいほど、定位精度が高いということになる。いずれの被験者においても、不特定モデルより適応モデルを用いた場合の方が、値が小さくなっていることが見て取れる。さらに、適応モデルを用いた場合、個人モデルを用いた場合とほぼ同等の値を得られていることから、定位位置の個人差を軽減できていることがわかる。

4 まとめ

本稿では、定位位置の個人差を軽減するために提案した個人適応手法の有効性を示した。今後は更なる定位精度の向上を図るため、本システムの利用中にオンライン個人適応を行なう手法について検討する。

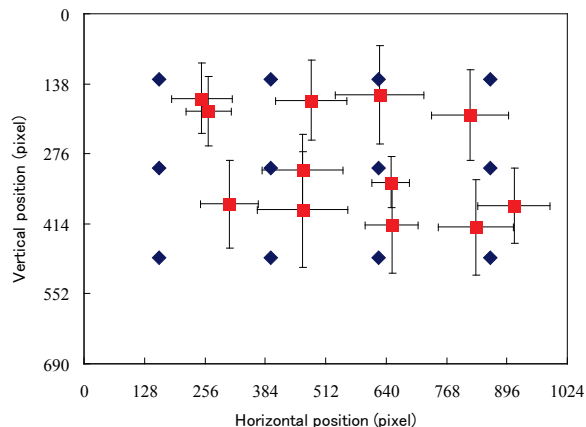


Fig.8: Localization result (independent model)

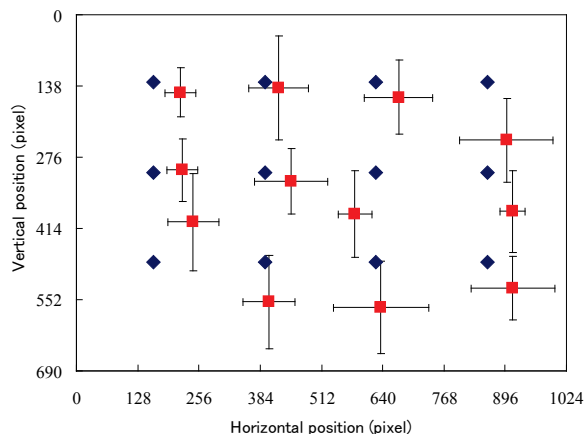


Fig.9: Localization result (adaptation model)

Table 3: Comparison of localization accuracy

	被験者 1	被験者 2	被験者 3
不特定モデル	97.7 pixel	145.1 pixel	90.3 pixel
適応モデル	75.5 pixel	103.3 pixel	76.7 pixel
個人モデル	74.4 pixel	98.1 pixel	61.2 pixel

参考文献

- [1] 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.
- [2] S. Amano, T. Yamada, S. Makino, and Nobuhiko Kitawaki, "A Study on Sound Image Control Method for Operational Support of Touch Panel Display," Proc. APSIPA ASC 2011, Thu-PM.PS2, Oct. 2011
- [3] 天野成祥, 山田武志, 牧野昭二, 北脇信彦, “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,” 日本音響学会秋季研究発表会, pp. 661-664, Sep. 2011.
- [4] 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,” 電子情報通信学会論文誌, Vol. 83-D-II, No. 12, pp. 2509-2516, Dec. 2000.