

ステレオステージ制御パラメータの決定法に関する考察

Consideration concerning decision method of stereo stage control parameter

武 晋太郎†
Shintaro Take

田中 章†
Akira Tanaka

河口 万由香†
Mayuka F. Kawaguchi

宮腰 政明†
Masaaki Miyakoshi

1 序論

多地点間のステレオ遠隔会議では、各地点が複数の遠隔地と接続されるため、そこに送られてくるステレオ信号は複数になる。このようなシステムでは、各地点で複数のステレオ信号を同時に再生する必要がある。会議参加者が各遠隔地のステレオ音像を容易に識別できるようにするためには、各地点のステレオステージ(ステレオ音像の創成領域)が互いに重ならないことが必要である。

2個のスピーカによる再生を前提としたステレオステージ制御方式としてこれまでにいくつかの手法が提案されている[1][2][3]。[1]は耳元への制御点での波形再現を目的としているので、受聴エリアは通常のステレオ再生より限定されたものとなる。[2]は原理的に音色変化を生じるという欠点を持っている。それに対して馬屋原らの提案した手法[3]は、ステレオ遠隔会議などの空間を想定した手法で、音色変化を抑えたままパラメータによってステレオステージ幅を制御することができる。また、単純な処理で構成され、現在のフレームの情報で処理を行うため、実時間で処理が可能である。しかし、マイクの位置などの情報を元にパラメータを設定しなければならないという問題がある。本論文では、[3]の手法に着目し、マイクの位置などの事前情報は仮定せず、入力信号から推定した到来時間差を用いて、ステレオステージの制御パラメータを自動的に決定する方法について考察する。

2 既存手法の概要

対象とするステレオステージ伸縮法[3]の手順について説明する。この手法は、周波数領域で振幅、位相を伸縮因子 κ を用いた行列 $\mathbf{K}_{\kappa, \Phi(\omega; \hat{t})}$ を用いて変換することでステレオステージの制御を実現する。以下、手順に沿って説明する。

まず、入力されたステレオ信号に短時間離散フーリエ変換を施し、離散スペクトル $\hat{X}_A^R(\omega; \hat{t})$, $\hat{X}_A^L(\omega; \hat{t})$ を得る。ここで、 ω は離散的な周波数を表し、 \hat{t} は窓の中心時間を表す。

次に、時間周波数点ごとに、スペクトル $\hat{X}_A^R(\omega; \hat{t})$, $\hat{X}_A^L(\omega; \hat{t})$ を、 $\hat{X}_B^R(\omega; \hat{t})$, $\hat{X}_B^L(\omega; \hat{t})$ に線形変換する。

$$\begin{pmatrix} \hat{X}_B^R(\omega; \hat{t}) \\ \hat{X}_B^L(\omega; \hat{t}) \end{pmatrix} = \mathbf{K}_{\kappa, \Phi(\omega; \hat{t})} \begin{pmatrix} \hat{X}_A^R(\omega; \hat{t}) \\ \hat{X}_A^L(\omega; \hat{t}) \end{pmatrix} \quad (1)$$

$\mathbf{K}_{\kappa, \Phi(\omega; \hat{t})}$ は線形変換の変換行列で、次式のように定義する。

$$\mathbf{K}_{\kappa, \Phi(\omega; \hat{t})} = \begin{pmatrix} ae^{-jb\phi(\omega; \hat{t})} & be^{ja\phi(\omega; \hat{t})} \\ be^{-ja\phi(\omega; \hat{t})} & ae^{jb\phi(\omega; \hat{t})} \end{pmatrix} \quad (2)$$

$$a = (1 + \kappa)/2 \quad (3)$$

$$b = (1 - \kappa)/2 \quad (4)$$

ここで、 $\Phi(\omega; \hat{t})$ は位相を表す周波数の実関数であり、 κ は伸縮因子である。位相 $\Phi(\omega; \hat{t})$ は次式のように決定する。

$$\Phi(\omega; \hat{t}) = \arg \hat{X}_A^R(\omega; \hat{t}) - \arg \hat{X}_A^L(\omega; \hat{t}) + 2\pi\psi(\omega; \hat{t}) \quad (5)$$

$\psi(\omega; \hat{t})$ は $\Phi(\omega; \hat{t})$ の範囲が $-\pi < \Phi(\omega; \hat{t}) \leq \pi$ となるような整数である。

最後に、逆離散フーリエ変換により、変換したスペクトル $\hat{X}_B^R(\omega; \hat{t})$, $\hat{X}_B^L(\omega; \hat{t})$ を時間領域信号に変換し、この短時間の信号に窓関数をかけてOverlap-Add[4][5]で時間軸上に配置することで、伸縮されたステレオ信号を得ることができる。

この手法ではステレオステージの伸縮を実現できるが、制御パラメータを決定する際にマイクの位置などの事前情報が必要となり、パラメータを適当に設定しなければならないという問題がある。

3 提案手法

本論文では図1のように、各地点からの複数のステレオ音像が重ならない範囲で広く位置するようなパラメータの値を自動的に求めることを目的とする。この際、マイクの位置などの事前情報は仮定せず、入力信号から算出した、到来時間差を用いる。

3.1 到来時間差推定

複数音源を含んだ左右チャンネルの入力信号を、 $-\alpha$ サンプルから α サンプルまで1サンプルずつずらして相関係数を求める。実際の到来時間差とずらしたサンプル数が一致するとき相関係数はピークを取るはずであるので、左右の最も大きいピークをステレオステージの幅とする。また、3音源以上が混合している場合は最も大きいピークは0付近に集まってしまうので、左右それぞれの初めのピークをステレオステージの幅とする。

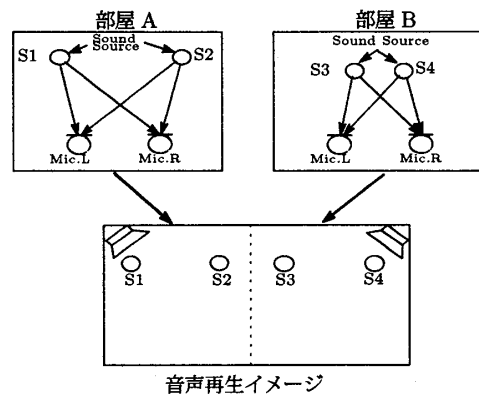


図1 ステレオステージ制御の概念図

† 北海道大学大学院情報科学研究科 CS 専攻

3.2 κ の決定

M 個の部屋からの複数音源が混ざっている入力音声を再生する際に、各ステレオステージを容易に識別するには次のような条件が必要であると考えられる。

1. 互いのステレオステージが重ならない
2. 部屋の広さを有効に使うために重ならない程度にステレオステージを拡張する

1. では、再生領域を $1/M$ に分割してそれぞれにシフトすることが必要となる。2. では、それぞれのステレオステージを一定の大きさにそろえる。これらを満たす κ を

$$\kappa = \frac{2s}{d_{right} + d_{left}} \cdot \frac{1}{M} \quad (6)$$

で与える。ここで、 d_{right}, d_{left} は入力信号から得た右、左の到来時間差で、 s は定数である。

式 (6) では、左右の到来時間差の和を s にそろえ、さらに $1/M$ に縮小する。到来時間差が大きいかほど音像は左右端に移動するが、最も端であるスピーカ位置から聞こえる時の到来時間差を s_{max} とすると、 $2m$ 離れて置いてあるスピーカから $2m$ 離れて聞いているとき、 $s_{max} = 28$ 程度で、1. の条件を満たすには $s = s_{max}/2$ が適当である。

また、左右の信号に遅延時間を与えることでステレオステージのシフトを行い、各ステレオステージは $1/M$ に分割した領域にそれぞれ割り当てる。

4 実験

実際の音声信号を用いて相関係数の計算により到来時間差の推定を行った。また、提案したパラメータ決定法に従って、音声信号の伸縮、シフトを行い、各ステレオステージが重ならず、広がっているかを検証した。

4.1 使用する信号

到来時間差の推定に用いた信号は、異なる文章を発話した複数個のモノラル音声信号である。これらの信号は、44.1[kHz] でサンプリングしたものである。この複数個の信号を図2のように到来時間差をつけて、複数音源を含むステレオ信号を作成した。また、相関を取る際は -50 サンプルから 50 サンプルまで 1 サンプルずつずらして求めた。

κ の決定法の検証実験では、 $M = 2$ の場合を想定して、2音源及び3音源が混合している信号を到来時間差を変えて作成した。到来時間差は $d = 7, 14, 21, 28$ [sample] の4種類を用い、聴き取りやすくするために各音源は再生時間が重ならないようにした。

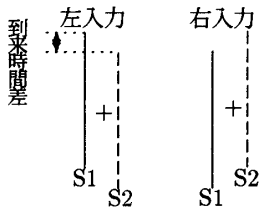


図2 モノラル信号 S1, S2 から作成した実験用ステレオ信号

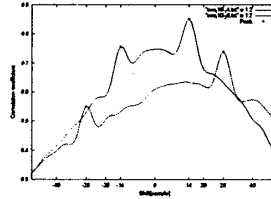


図3 2音源が混合した信号から計算した相関係数

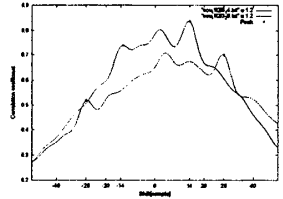


図4 3音源が混合した信号から計算した相関係数

4.2 実験結果

2音源及び3音源が混合した信号を用いた到来時間差推定の結果を図3,4に示す。2音源では2つ、3音源では3つのピークが現れていることが確認でき、どちらにおいても到来時間差 = 14, 28[sample] が推定された。これにより両端のピークを左右の到来時間差であると推定することができる。さらに到来時間差を 7, 14, 21, 28[sample] として、この到来時間差から (6) で決定したパラメータを用いて聴取実験を行った。この結果、どの到来時間差に対してもステレオステージの分離が主観的に確認でき、 $d = 7$ の場合は拡大、 $d = 28$ の場合は縮小していることが確認できた。

5 結論

本論文では馬屋原らのステレオステージ制御法に対して、マイクの位置などの事前情報は用いずに左右の信号の相関から推定した到来時間差を用いて、ステレオステージの制御パラメータを自動的に決定する手法を提案した。今後の課題としては、精度の良い到来時間差の推定法、レベル差を考慮した再生法などが挙げられる。また、音像の再生位置には視覚情報も影響するので [6]、視覚情報による定位位置を考慮に入れた再生法も今後の課題である。

参考文献

- [1] Ronald M. Aarts, "Phantom Sources Applied to Stereo-Base Widening," *J.Audio Eng. Soc.*, Vol.48, No.3, March 2000.
- [2] Michael A. Gerzon, "Applications of Blumlein Shuffling to Stereo Microphone Techniques," *J.Audio Eng. Soc.*, Vol.42, No.6, June 1994.
- [3] 馬屋原立裕, 穂刈治英, 島田正治, "スペクトル補間を用いたステレオステージ制御方式-他地点通信のための新しい要素技術-", 信学論 A, Vol.J85-A, No.8, pp.819-832, Aug. 2002.
- [4] Daniel W. Griffin, Jae S. Lim, "Signal Estimation from Modified Short-Time Fourier Transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.ASSP-32, No.2, April 1984.
- [5] Jont B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.ASSP-25, No.3, June 1977.
- [6] 樋渡涓二, "視聴覚情報概論," 昭晃堂, 1986.