

[ポスター講演] 音のパンニング加工の検出に関する一検討

黒田 康弘[†] 藺田光太郎[†] 喜安 千弥[†]

[†] 長崎大学工学部工学科情報工学コース

〒 852-8521 長崎県長崎市文教町 4-11

E-mail: †bb35315017@ms.nagasaki-u.ac.jp, ††{sonoda,kiyasu}@cis.nagasaki-u.ac.jp

あらまし 本稿では、音データに含まれる人工的な加工の程度を定量化し評価できるシステムを提案する。近年では、音を模擬する技術が高まり人間は本物の音と人工的な模擬音との弁別が難しくなっていることが問題となっている。そこで、計算機を用いて人工的な加工の程度を提示することで模擬音によるシステムへの攻撃を防ぎ、音楽の新たな価値として利用することができると考える。本稿では、デジタル音響処理の一部であるパンニング処理を用いて加工したステレオ音データを用いて、音データのパワー比、群遅延差を用いて加工の程度を測るシステムを提案する。

キーワード 擬似ステレオ化, 音のパンニング加工

[Poster Presentation] A study on detection of panning processing of sound

Yasuhiro KURODA[†], Kotaro SONODA[†], and Senya KIYASU[†]

[†] Computer and information science program, Department of engineering, Nagasaki university

4-11 Bunkyo-machi, Nagasaki-shi, Nagasaki 852-8521, Japan

E-mail: †bb35315017@ms.nagasaki-u.ac.jp, ††{sonoda,kiyasu}@cis.nagasaki-u.ac.jp

Abstract In this paper, we propose a system that can quantify and evaluate the grade of liveness of sound data. In recent years, technology to simulate sound has improved and it is difficult to distinguish real sounds from artificial simulated sounds by human hearing system. Therefore, by using a computer to notice the extent of artificial processing, it make us prevent the attack on the system by simulated sound and use it as new worth of music. In this paper, we propose a system to measure the grade of liveness against artificial stereo panning processing by using deviations of power ratio and group delay difference between L-R channels of sound data.

Key words Pseudo-stereo, Panning potentiometer

1. ま え が き

現代の音楽制作現場では、コンピュータの利用により簡単にデジタル音響加工を行うことが可能になった。一方で、人工知能の導入により、簡単に人間には聞き分けることが出来ない高品質な模擬音を作成できてしまうことも問題視されている。例えば、工場の異常音を模擬した音が作成され、再生された場合、労働者に誤った判断をさせてしまう。また本物であることに価値がある音源データなどを模擬されることがあげられる。そこでデジタル加工の程度を人間に代わってコンピュータに検知させ、それを人間に提示することで、模擬音によるシステムへの攻撃に対して頑健性を持つことができると考えられる。

今回はデジタル音響処理の一つである音のパンニング加工を検出し、その程度を評価することを目標とした。パンニング処理とは、モノラル音に対して音の定位を人工的に変化させたステレオ音を構成することにより空間的な演出を行うことが出

来る音響加工である。実際に角度をつけて音源を定位させて録音した音源と、人工的に定位するようにパンニング加工を施した音源とを比較し、どちらが加工された音源であるか人間は聞き分けることは難しい。そこで、L と R において音のパワー比および群遅延差の時系列の標準偏差から人工的なパンニング加工を検出、程度を数値として評価した。

2. 音のライブネス度の定義

本稿では便宜上のため「音のライブネス度」という尺度を新たに定義し説明する。ライブネスとは、英単語の"Live"から取った言葉であり、音源にデジタル音響処理をかけていない自然のものであることをさす。また、音源に含まれる人工的な加工の低さの度合いを「ライブネス度」と定義した。ライブネス度はデジタル音響処理を用いて作られた模擬音の検出に役立てたり、生の楽器の音源データであることを保証する指標することを目標とする。

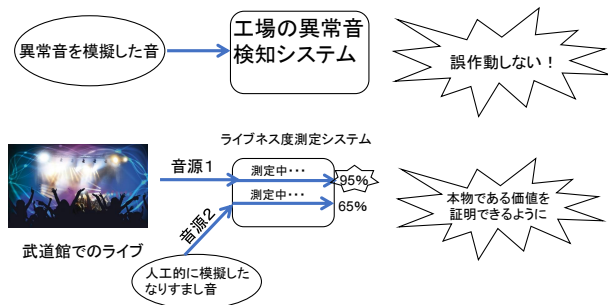


図1 ライブネス度システムの例

関連研究では、矢口ら [1] は、スピーカ再生による音声のリブレイアタックを検出している。音声収録中の無発話区間において、スピーカ再生では暗騒音のほかに周期的なスピーカ雑音が残ることを利用した手法である。この手法は話者照合認証における生体検知 [2] に使用される。一方で音楽はむしろリプレイされる音楽を楽しむことが多く、本研究では、再生されるステレオ音源がライブ音源なのか人工的に加工された音源なのかを対象とする。そこで、音声だけに限らず音楽など一般的な音において人工的な特徴（加工）の検出を可能にし、図1のように音源をライブネス測定システムに通すことでライブネス度を提示するようなシステムの構築を考える。本稿では、CD音源の編集に使われているデジタル音響処理の中で音楽制作現場での擬似ステレオ技術の一つであるパンニング加工に焦点を絞ってその加工度合いを評価した。

3. パンニング加工検出

この章では、音データに含まれるパンニング処理を検出し評価する手法について記述する。

3.1 パンニング処理について

パンニングとは、パンポット (panning potentiometer) が由来となっており、単一のマイクロホンで収録した音源の信号をステレオの左右チャンネルに振り分けて左右スピーカ間の任意の方向に定位させる装置のことであり、ソフトウェア上で同様の処理を行うことをパンニングと呼称する [3] [4]。パンニング処理は図2や図3のイメージのように、人間が聴くうえで左右のスピーカに音量差や時間差をつけることで聴感上の音源の定位の位置を変化させる。

パンニングの方式には、音圧パンニングと時間差パンニングの二種類が存在する。音圧パンニングとは、図2のように1ch信号に対して定位角度によって定められるL,Rそれぞれのパンニングカーブの値をゲインとして、L,Rの2ch信号を構成する処理である。時間差パンニングは、図3のように1ch信号に対して定位角度によって定められる位相差を付けたL,Rの2ch信号を構成する処理である。なお今回パンニング処理を行うにあたって、音楽制作の現場で一般的に使われているDAW(Digital Audio Workstation) ソフトであるCubase9 [5]を使用した。

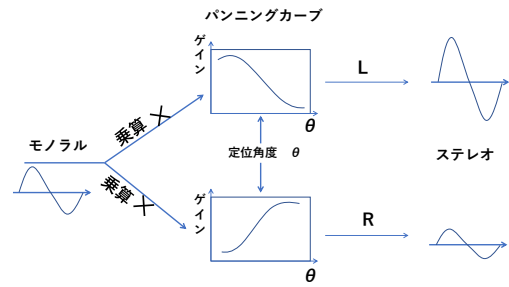


図2 音圧パンニング

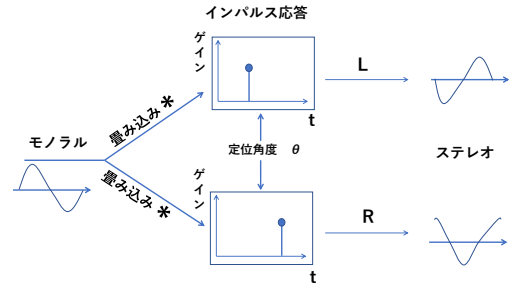


図3 時間差パンニング

3.2 提案手法について

(1) 音圧パンニングではL,Rの位相差はつけない。したがってLとRのパワー比 IPR(Inter-channel Power Ratio) を考える。

$$IPR = \frac{P_L}{P_R} = \text{const} \quad (1)$$

式(1)で示すように、人工的な音圧パンニングを用いた場合はL,Rのパワー比が時系列上の短い区間では一定になると考えられる。

(2) 時間差パンニングでは、常にLとRにつけられた時間差が一定である。したがって、群遅延差 IGDD(Inter-channel Group Delay Difference) を考える。群遅延を式(2)で示す。

$$G(\omega) = -\frac{d}{d\omega} \phi(\omega) \quad (2)$$

L,Rにおける群遅延差を式(3)で示す。

$$IGDD = G_L(\omega) - G_R(\omega) = \text{const} \quad (3)$$

式(3)で示すように、人工的な時間差パンニングを用いた場合はIGDDが時系列上の短い区間では一定になると考えられる。

したがって、人工的なパンニング加工を行った音源では式(1)や式(3)のように、IPR,IGDDは一定になりやすく、対して生音では定位が揺らぎやすいため、人工的なパンニング加工を行った音源と比べて、IPR,IGDDは時系列上でばらつきやすいと考えられる。そこで、パンニングによって人工的に定位させた音源のIPRとIGDDの時系列変化は、生音にくらべてばらつきが少ないという仮説を立て実験を行った。提案手法を以下に示す。

提案手法

(1) 2ch の時間信号 $v(t)$ に一定のフレーム長で STFT を行い、 $V(\omega)$ のパワースペクトル $P_L(\omega)$, $P_R(\omega)$ と位相スペクトル $\phi_L(\omega)$, $\phi_R(\omega)$ を求める。

(2) 左右のパワースペクトル $P_L(\omega)$, $P_R(\omega)$ の比の全 ω の平均を $IPR(t)$ に格納する。

(3) 位相スペクトル $\phi(\omega)$ は群遅延 $G(\omega)$ を取ったのち L,R の群遅延差の全 ω の平均を $IGDD(t)$ に格納する。

(4) すべての t において、1 から 3 の手順を繰り返す。

(5) $IPR(t)$ と $IGDD(t)$ において、フレーム数 32 個ずつのブロックに区切っていきそれぞれの標準偏差を短区間音圧ライブネス度、短区間群遅延ライブネス度とし、全ブロックのそれぞれの平均を平均音圧ライブネス度、平均群遅延ライブネス度と定義する。

今回設定したフレーム長は 1024 である。また IPR は対数変換している。今回の評価には IPR と $IGDD$ における、平均音圧ライブネス度、平均群遅延ライブネス度の値を使用して評価する。

4. 提案手法の評価と考察

今回実験にて使用した音源は、図 4 のように収録した。Mic_L と Mic_R はアコースティックギターを正面 0 度、45 度に見るように 2m 離れた位置に配置した。Mic_L, Mic_R はすべて同種の無指向性マイクである。Mic_L, Mic_R からステレオ音源 ("自然音") を作成、Mic_L を R 方向に 31 %, 75 %, 95 % のパンニング加工を行ったステレオ音源をそれぞれ作成した。使用音源を表 1 に示す。

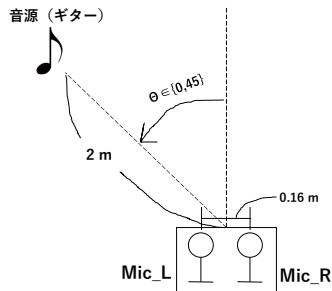


図 4 録音実験図

表 1 使用音源

音源名	内容	加工
自然音	L,R 二本の無指向性マイクで収録した生音 (0 度)	
自然音 2	L,R 二本の無指向性マイクで収録した生音 (45 度)	
pan31	Mic_L の音源を L 方向 31 % のパンニング加工	あり
pan75	Mic_L の音源を L 方向 75 % のパンニング加工	あり
pan95	Mic_L の音源を L 方向 95 % のパンニング加工	あり

4.1 音源データの評価

まず音源 "自然音" と音源 "pan31" の音圧比 IPR 、群遅延差

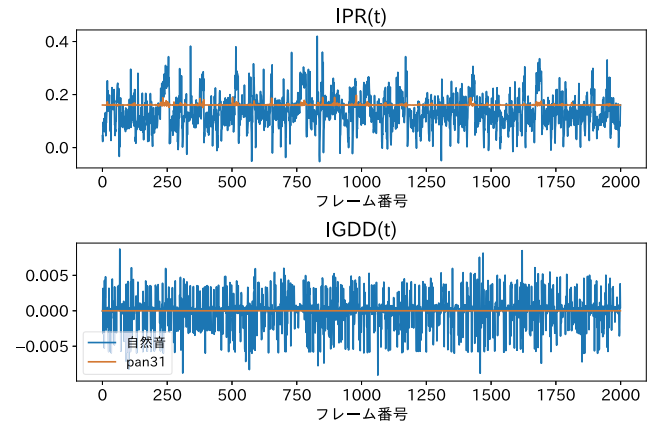


図 5 自然音 (青) とパンニング加工した音 (オレンジ) の比較

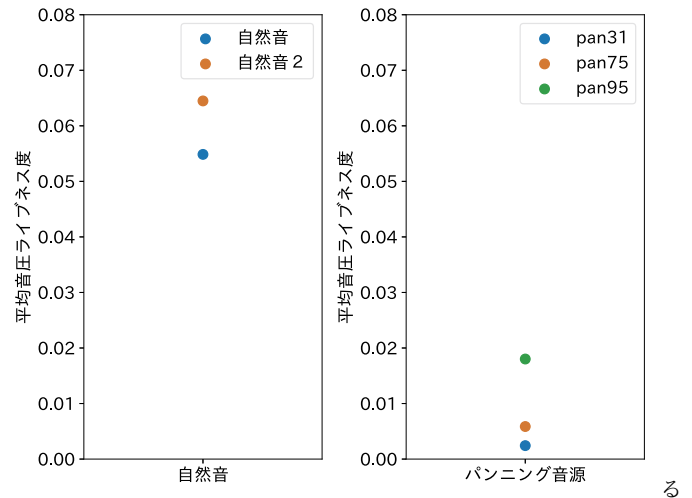


図 6 音源データの平均音圧ライブネス度の比較

$IGDD$ の変化を図 5 に示す。図 5 より、一定のフレームごとの L,R の音圧比および群位相差は自然音のほうが大きくばらついた分布をしており、人工的にパンニング加工を行った音源は一定に近いことが見て取れる。

音源 "自然音", "自然音 2", "pan31", "pan75", "pan95" における平均音圧ライブネス度を図 6 で示す。人工的なパンニング加工を行った音源は自然音と比べて平均音圧ライブネス度の値は小さい。また、同じように平均群位相ライブネス度を図 7 に示す。同様にパンニング加工を行った音源は自然音と比べて平均のライブネス度は低い。

4.2 楽曲データの評価

次に、生音を録音したライブネスな楽曲と、個別に複数の楽器を録音し適切にマスタリングされた楽曲の二種類を比較した。

さらに、実験用の音楽データとして RWC 研究用音楽データベースより、ポピュラー音楽・著作権切れ音楽データベースから 3 曲 [8]、クラシック・ジャズ音楽データベースから 3 曲 [7] を選択した。サンプリング周波数は 44.1kHz で録音、WAVE ファイルで保存し、角度は正面に対して R 方向、L 方向である。

RMC 研究用データベースに収録されている楽曲は、ProTools [6] という DAW ソフトを用いて、すべての録音、トラッ

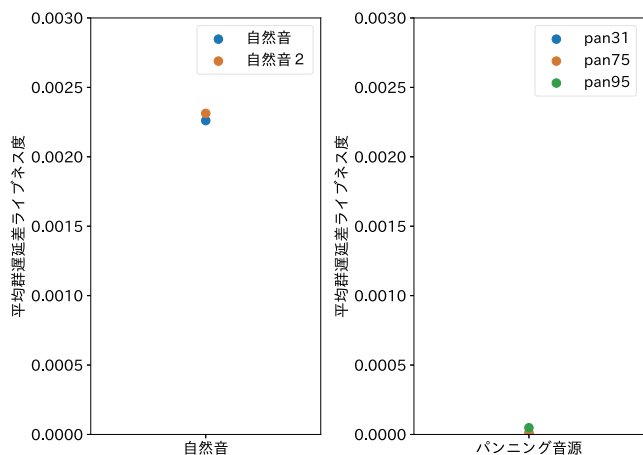


図 7 音源データの平均群遅延ライブネス度の比較

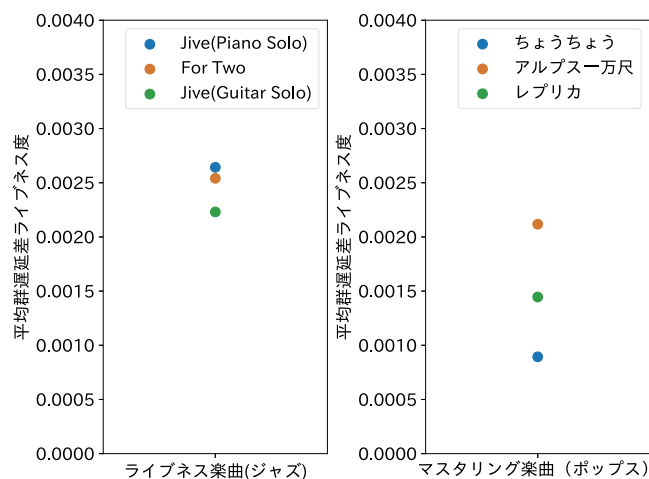


図 9 "楽曲における平均群遅延ライブネス度の比較"

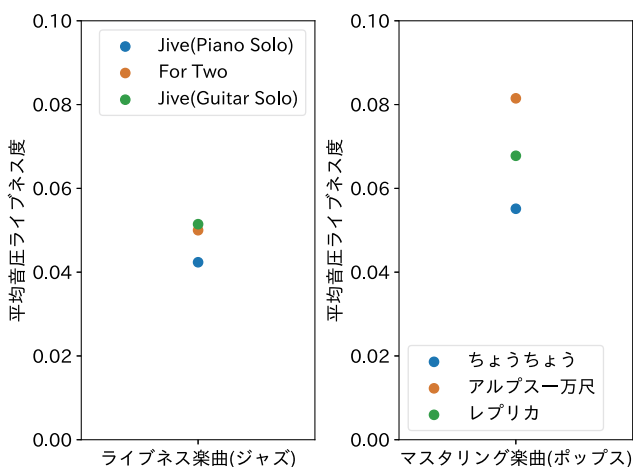


図 8 楽曲における平均音圧ライブネス度の比較

クダウン、トータルエフェクト、マスタリングを行っている。通常、ポピュラー楽曲では、楽器ごとに個別収録することが多く、必然的に別トラックに録音される。一方で、クラシック音楽やジャズ音楽では、より良い演奏や場のニュアンスを実現する目的で、同じ空間で同時に演奏して収録することが多い。今回実験で使用するポピュラー・著作権切れ音楽データベースの楽曲は人工的なパンニング加工を含んでいるライブネス度の低い楽曲であり、対してクラシック・ジャズ音楽データベースは生演奏をそのまま録音したことに近くライブネス度の高い楽曲と考えられる。

楽曲における平均音圧ライブネス度を図 8、平均群位相ライブネス度を図 9 に示す。図 8 より、生音を録音したライブネスな楽曲は、人工的にマスタリングした音源と比べて平均音圧ライブネス度が低い結果となった。対して、図 9 より、ライブネスな楽曲は、人工的にマスタリングした音源と比べて平均群遅延ライブネス度は高い結果となった。したがって、平均音圧ライブネス度では人工的な加工を検知できていないとは言えず、平均群遅延ライブネス度では正しく人工的な加工を検知できていると考察される。

5. 結 論

本稿で提案したシステムでは、パンニング加工を行った音源は生音と比べライブネス度は低い値となり、L,R のパワー比と群遅延差を用いたライブネス度を測ることで弁別可能であることがわかった。今後の課題としては、パンニング加工以外のデジタル音響加工をかけた音源に対してライブネス度を計る手法を新しく確立する必要があると考えられる。例えば、ドラムのスネアの音などは、極めて短時間の音程感のない音を発するためパワーや位相に大きな変化や特徴が表れにくく、パワー比、群遅延差を用いた本手法ではうまくライブネス度を評価することが難しいと考えられる。

文 献

- [1] 矢口 凌也, 塩田 さやか, 小野 順貴, 貴家 仁志, "複数チャネル間の相互相関関数を用いた話者照合のためのなりすまし検出," 日本音響学会 2018 年秋季研究発表会講演論文集, pp.1335-1338, 2018 年.
- [2] 山岸純一, "ASVspoof:話者照合における生体検知," 日本音響学会秋季研究発表会講演論文集, pp.1329-1330, 2018 年.
- [3] 青木直史, "C 言語ではじめる音のプログラミング サウンドエフェクトの信号処理 第9章 音を広げる," オーム社, pp.126-129.
- [4] 日本音響学会 "新版:音響用語辞書," コロナ社 pp.317.
- [5] 音楽制作ソフトウェア, "Cubase", <https://new.steinberg.net/ja/cubase/>
- [6] 音楽制作ソフトウェア, "Pro Tools", <https://www.avid.com/ja/pro-tools>
- [7] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, "RWC 研究用音楽データベース:クラシック音楽データベースとジャズ音楽データベース", 情報処理学会 音楽情報科学研究会, 2001-MUS-44-5, Vol.2002, No.14, pp.25-32, February 2002.
- [8] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, "RWC 研究用音楽データベース:ポピュラー音楽データベースと著作権切れ音楽データベース", 情報処理学会 音楽情報科学研究会, 2001-MUS-42-6, Vol.2001, No.103, pp.35-42, October 2001.