

# 携帯型音楽プレイヤーのための雑音・音楽融合型再生方式の評価

井上 亮文<sup>†</sup> 備瀬 翔平<sup>†</sup>  
市村 哲<sup>†</sup> 松下 温<sup>††</sup>

携帯型オーディオプレイヤーの普及が目覚ましいが、屋外の試聴でユーザを悩ませる問題の1つとして雑音の存在がある。本論文では、楽器演奏の知識を利用して周囲の雑音を感覚的に低減させる手法を提案する。提案手法では「リズム楽器のアクセントが一致した演奏は聴いて心地良い」という経験則を利用し、周期性のある雑音をドラムやベースなどの楽器演奏に見立てる。そしてそのビートに音楽のビートが一致するよう音楽の再生速度をユーザが知覚できない程度に変換して再生する。電車の走行音を用いた検証により、雑音を信号レベルで低減させなくても聴きやすさが向上したこと、音楽に無理な改変を加えなくても一定の効果が得られたことを確認した。

## Evaluation of Music-noise Assimilation Playback for Portable Audio Players

AKIFUMI INOUE,<sup>†</sup> SHOHEI BISE,<sup>†</sup> SATOSHI ICHIMURA<sup>†</sup>  
and YUTAKA MATSUSHITA<sup>††</sup>

Though the population of portable audio listeners keeps growing, it still carries issues such as distractive background noise. This paper proposes a background noise reduction method using knowledge obtained by musical instrument performances. The method is based on our experience that listening to music feels comfortable when the rhythm accents are matched up at the correct timing. In our method, noise with a periodic rhythm takes the role of the rhythm section such as drums and basses. Then the tempo of the current track is converted at an unnoticeable amount so that the accents of the track can synchronize with that of the noise. We examined the reduction effects by carrying out five experiments, in which the traveling sound of a train was set as the periodical noise. Results of the experiments showed that the noise was faded out to a pleasant level without carrying out any spectrum noise reduction. We also showed that unreasonable changes in the tempo were unnecessary to keep a constant effectiveness.

### 1. はじめに

携帯型デジタル音楽プレイヤーの普及が目覚ましい。わずか数十グラムの重量で膨大な数の曲を持ち運べ、しかも長時間再生可能であることから、多くのユーザが屋内外問わず音楽を楽しんでいる姿を頻繁に見かけるようになった。このようないつでもどこでもの音楽鑑賞は、もはやライフスタイルの一部になったといえる。

しかし、産業革命以降の300年で生活の中での音が急増した<sup>1)</sup>。特に屋外では、音楽の視聴を妨害するよ

うな不必要な音(雑音)に遭遇することが多い。20数年前のプレイヤーと比べてサイズ、音質、バッテリー駆動時間など多くの面で性能が改善されたとはいえ、この雑音にどう対処するかは相変わらずの課題として残っている。

代表的なノイズ対策の1つは密閉型ヘッドフォンである。これはイヤークッションで外耳ごと密閉し、ハウジング内の空気を利用して外部からの音を遮蔽することができる。もう1つの例はノイズキャンセリングヘッドフォンである。これはヘッドフォンに内蔵されたマイクロホンが周囲の音を拾い、騒音に反対に作用する逆位相の音をノイズキャンセリング回路で作る。そしてその逆位相の音を重ね合わせることにより、外部の音が打ち消される仕組みを持つ。

これら既存の対処法は、いずれも空気の振動という音の物理的な特徴を利用し、雑音を内耳に到達しない

<sup>†</sup> 東京工科大学コンピュータサイエンス学部  
School of Computer Science, Tokyo University of Technology

<sup>††</sup> 住宅情報化推進協議会  
ALICE FORUM

ようにすることで音楽だけに閉じこもる空間を作り出している。しかし、遮音性を高めすぎると必要な音も聞こえなくなるため、これらの製品は安全性の観点からも一定の外部音が聞こえるようにしている。その音が気になるユーザは再生音量を大きく設定してしまい、音漏れにより周囲に迷惑をかけたり、聴力障害の原因になったりしている。

本研究では、楽器演奏の知識を利用して周囲の雑音を感覚的に「気にならなく」させる手法を提案する。この手法では、雑音の存在を許容しつつ音楽の中うまく溶け込ませ、音楽そのものを聴きやすくすることを目的とする。この実現のため、雑音をリズム楽器の演奏に見立ててそのテンポを測定する。そして雑音のリズムと音楽のリズムがぴったり一致するように、音楽の速度をユーザが感知しない程度に変更して再生をする。本論文では、雑音を電車の走行音、音楽はポップスなどの現代音楽と限定したうえで、その効果の検証評価を議論する。

本論文の構成を以下に示す。2章では現代音楽の特徴と人間の聴覚的特性について、3章では提案手法について、4章では実装について、5章では評価実験について述べ、6章では実験結果をふまえてシステムを実現するうえでの課題について議論し、7章をまとめとする。

## 2. 現代音楽と人間の特性

本章では、音楽と雑音を融合させるうえで注目した現代音楽の要素と、人間の聴覚・感覚的特性について述べる。

### 2.1 ビート

現代音楽を特徴付けるものの1つに、ビートの存在がある。ビートは小節内における単位時間の区切りを表し、一般的にリズム楽器が音を鳴らす箇所である。

拍数が奇数の部分を強拍、偶数の部分を弱拍と呼ぶ。音楽を演奏する際は、バスドラムとスネアドラムを同時にたたくのは避け、強拍をバスドラム、弱拍をスネアドラムという形で使い分けることが多い。したがって、通常のポップスやロックでは、1, 3, 5, 7拍目にビートを置くことが多い。

### 2.2 感覚的特性

音楽を演奏するうえで、特にベースやバスドラムといったリズム楽器を演奏する際は、互いのビートを一致させるのが鉄則である。現代音楽を聴きなれたユーザは、このアクセントが一致していると一体感を覚え、心地良く感じる。逆にずれていると、なんとなく散漫な印象となる。いわゆる「ノリが悪い」状態である。

「beatmania」や「太鼓の達人」に代表される、いわゆる「音ゲー」は、この感覚的特性を利用したものである。これらゲームは音楽に合わせてステップを踏んだり、楽器を模したコントローラを使って演奏をしたりする。そしてプレイヤーの操作が音楽のビートと一致すれば得点上がり、逆に一致しなければ減点されていく。まさしくビートの一致を利用して心地良さを味わうものであり、これらが大衆に広く受け入れられていることは、音楽を楽しむうえでのビートの重要性を如実に表している。

## 3. 提案手法

### 3.1 理論

雑音の影響が大きいときにユーザが再生音量を大きくするのは、人間の聴覚特性の1つのマスキングに関連する。これは、同時に聞こえる音の大小によって聞き取れる音圧レベルの最小値が変化するという特性を利用し、所望信号である音楽の音圧レベルを大きくして雑音を知覚させにくくする操作であるといえる。

しかし、現代音楽は1つの音が鳴り続けているわけではない。文章において節や文が存在するように、音楽では複数の音が、ある順序、あるタイミングで鳴ることによって系列を作り、意味を持ったまとまり「フレーズ」を形成している(図1(a))<sup>2)</sup>。よって音楽中の演奏音に類似した雑音が入ると、それぞれの音が音圧レベルでマスキングされるだけでなく、意味を持った音系列全体としても影響を受けると考えられる。たとえば図1(b)では、元のメロディラインとは異なる雑音が入っている様子を表している。この結果、音楽を聴いているユーザにとっては、音数が増えて感覚的な速さ(リズム感)が変わったり、現在のコードと親和性のない音が入って不安定な響きになったりする。

本研究で提案する理論は、この意味レベルへの影響を最小限にとどめようという発想に基づいている。図1(c)のように、音楽の意味的な構造を理解して音系列の間に雑音が入らないように音楽を再生することができれば、影響は重なった音だけにとどめられ、メロディラインの大部分は維持される。さらに、重なる音の特徴が類似している場合には、雑音を本来の音楽の一部として解釈するのではないかと考えられる。

### 3.2 基本方針

本節では、前節の理論を実現するうえでの具体的な

1997年にコナミより発売。ターンテーブルを模したコントローラを使い、クラブDJになりきるゲーム。

2001年にナムコより発売。パチを模したコントローラで太鼓奏者になりきるゲーム。



図 1 雑音がメロディに与える影響  
Fig. 1 Noise effects on melody line.

方針について述べる。

まず、どのような雑音を対象とするかを議論する。本研究では雑音を音楽中の特定の音と重ね合わせるため、雑音の発音位置があらかじめ予測できるものが望ましい。実際、音楽の視聴時に問題となる雑音の中には、電車や車がレールや道路の継ぎ目を走行する際の「ガタンゴトン」という音のように、一定の周期で類似パターンが出現するものも多い。そこで本研究では、等速運行中の電車の走行音を対象とする。

次に、元の音楽のどこに雑音を重ねればよいのかを議論する。電車の走行音は、特定の音階を表すというより、一定のテンポで演奏されるリズム楽器のようなものと考えることができる。2章で述べたように、リズム楽器のアクセント位置であるビートの一致・不一致は、現代音楽を視聴するうえで非常に重要な役割を果たす。3.1節の理論から見れば、音楽の視聴中に軌道走行音が気になるのは、双方のビート位置が一致せず意味レベルの影響、特にリズム感の受けていることが一因と考えられる。よってこれを一致させれば、ユーザの不快感は雑音を低減させなくても感覚的に減少する可能性がある。

最後に、どのようにしてビート位置を一致させればよいのかについて議論する。プレイヤーの再生位置を遅らせるなどして音楽と雑音のビートをある時点で一致させたとしても、両者のテンポ (BPM: Beat Per Minute) に差があればすぐに一致なくなってしまう。BPM一致のためには、(1) 雑音の速度を変える、(2) 音楽の速度を変える、の2通りが考えられるが、雑音の速度を変えるには電車の走行速度を変える必要がある。本研究では電車の一乗客であるユーザが音楽を聴くことを想定しているため、前者は難しい。よって音楽の速度をユーザに知覚されない程度に変更する

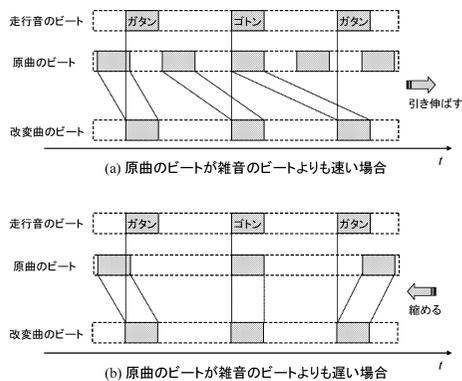


図 2 提案手法の概要  
Fig. 2 Overview of the proposed method.

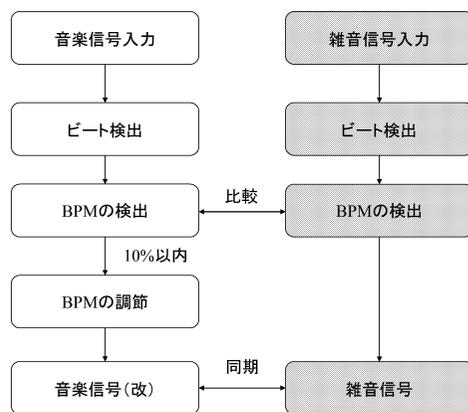


図 3 提案手法の処理の流れ  
Fig. 3 System flow.

立場をとることとする。

図 2 に提案手法の概要を示す。横軸は時間の流れ、網掛けの矩形部分は雑音、原曲それぞれのビート位置を表している。図 2(a) のように原曲の BPM が雑音に比べて若干速い場合は、時間軸を引き伸ばして音楽の BPM を遅くすることでビート位置を一致させる。逆に図 2(b) のように原曲の BPM が遅い場合は、時間軸を縮めて速くすることで一致させる。

### 3.3 処理の流れ

図 3 に提案手法の処理の流れを示す。まず、音楽・雑音双方の信号を入力し、それぞれに対しビート位置の検出を行う。次に、BPM の検出を行い、双方の差を比較する。速度に大きな差がない場合は、双方の BPM が一致するよう音楽の BPM 変換を行う。最後に、BPM が変換された音楽と雑音のビート位置を同期させ再生を行う。

なお、本手法では音楽の選曲にはいっさい関与しない。基本的にユーザが聴きたい曲を聴きやすくすることとする。聴きやすさの向上度合いは、雑音と音楽の

BPM や周波数特性によって曲ごとに異なると考えられる。しかし、最も聴きやすい曲が必ずしもユーザが望んでいる曲とは限らないからである。

### 3.4 速度変換

音楽コンテンツの速度変換をする場合、単純に時間軸を操作しただけでは音程が変わってしまう。一般に、人間の周波数弁別閾は 250 ~ 4,000 kHz で 0.2% 程度といわれており、音程に関して非常に敏感である。携帯型プレイヤーのユーザは同じ曲を日々繰り返し聴いていることが多いため、そのような変換を行うと原曲を聴きなれているユーザは違和感を覚えてしまう。

そこで本手法では、曲の音程を保ったまま時間軸を操作する TSM (Time Scale Modification) 処理を施す。人間は、音程に変化がない場合、音楽の再生速度が多少変化しても気がつきにくいという特性がある。実際にこの操作は、BPM の異なる 2 曲をスムーズに連続再生 (ミキシング) する DJ (Disk Jockey) が多用するテクニックでもある。

しかし、変換の前後で速度があまりにも違いすぎると曲の印象が変わってしまう。DJ へのヒアリング調査によれば、ミキシングする際の変換は数パーセントが限界という。実際の楽曲分析の結果からもその傾向は顕著である<sup>3)</sup>。本研究ではこの結果を参考に、提案手法を適用する限界を  $\pm 10\%$  と設定した。

## 4. 実装

提案手法の効果を検証するための実験用プレイヤーを作成した。実装は Windows 上で Visual C++ で行い、雑音の仮想ビートに同期して音楽 (WAVE, 44.1 kHz, 16 bit, ステレオ) を再生する。このプレイヤーは、1 つの音楽ファイルと、1 つの雑音が記録されたファイルを入力として受け取り、図 3 に示した流れで処理を進める。最終的に、雑音とビート位置を同期させた音楽ファイルを出力する。

雑音として用いる電車の走行音は、実際に録音したものをループ編集して繰り返し用いた。本来は雑音をマイクで拾い、それにリアルタイムで同期できることが望ましいが、本論文では提案手法の効果検証に重点を置いている。よって実験を容易にするため再現性がある形式にした。

雑音のビート検出は、まず図 4 に示す波形抽出用のハイパスフィルタ関数を定義した。次に周波数領域でこのフィルタ関数をかけたあと、時間波形のピーク閾値処理を行うことで実現した。フィルタ処理後の波形 (約 4 秒) を図 5 に示す。丸印のついている部分がビートと判定した位置である。なお、この結果得られ

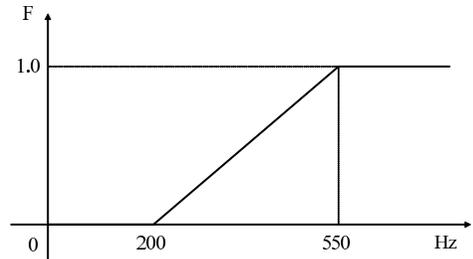


図 4 雑音に対するハイパスフィルタ関数  
Fig. 4 HPF for train noise.

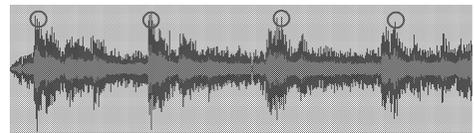


図 5 電車走行音のフィルタ処理後の波形  
Fig. 5 The waveform of filtered noise.

表 1 SOLA 用パラメータ  
Table 1 Parameters for SOLA.

パラメータ名	値
$\alpha$	$0.9 < \alpha < 1.1$
$S_a$	$2,000 / \alpha$
$S_s$	2,000
$k_{min}$	-400
$k_{max}$	400
$N$	2,800

た雑音の BPM は 65 であった。

音楽のビート検出は、事前に手動で測定したものをを用いた。ビート位置やリズム楽器の発音位置は自動で検出する方法<sup>4),5)</sup>も提案されており、いくつかの手法はソースも公開されている。しかし正確に検出できない場合があると試聴評価に影響が出るため、今回は事前に計測した BPM 値とビート位置を利用することとした。

曲の音程を一定に保ったまま速度を変更する TSM は、SOLA (Synchronous Overlap and Add) 法<sup>6)</sup>を用いた。提案手法は外部の雑音に応じてリアルタイムに曲の伸縮を行う必要がある。SOLA は他の手法に比べて音質面では多少劣るものの、アルゴリズムが比較的単純で処理が軽いという特徴がある。

文献 7) で定義された OLA および SOLA に必要なパラメータである速度比  $\alpha$ 、解析フレーム間隔  $S_a$ 、合成フレーム間隔  $S_s$ 、アライメント調整範囲  $k_{min}$  および  $k_{max}$ 、分解フレーム長  $N$  は、文献 8) をもとに表 1 のように決定した。このうち  $\alpha$  の限界は 3.4 節で述べた値を用いた。

図 6 にビートを同期させた場合の振幅を Screen-

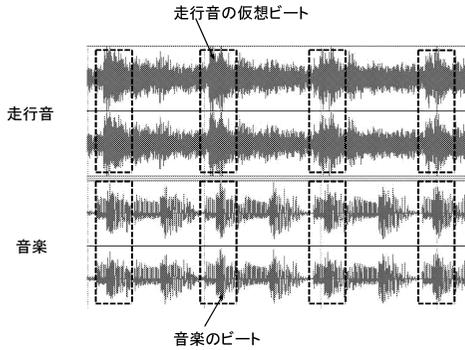


図 6 ビート同期後の振幅の様子

Fig. 6 Amplitudes of music and noise after the synchronization.

blast Movie Studio で表示した例を示す。上が雑音，下が BPM 調整を行った音楽である。双方のピーク位置がほぼ一致しているのが分かる。

### 5. 評価

#### 5.1 実験方法

本研究で提案，構築を行ったシステムを使用して，20 人の被験者を対象に評価実験を行った。

本研究の目的は，マスキングにより雑音を聴こえないようにすることではない。雑音が知覚できる状態でも，音楽としての意味系列が維持された「気にならない」状態を作り出すことである。よって実験では「聴こえる/聴こえない」というマスキング量は扱わず，雑音が聴こえる状態で「気になる/気にならない」を主観的に評価してもらうことを目的に，以下に述べる 5 種類の実験を用意した。

**準備** 実験に用いた雑音と音楽中ビートの周波数特性を図 7 に示す。音楽は邦楽のポップスであり，ボーカル中心で演奏楽器数が少なく，混入した雑音が目立ちやすいものを選んだ。ビート音はベースと手拍子であり，ベースは 150 Hz，手拍子は 2,000~4,000 Hz 近辺，走行音は 200 Hz~2,000 Hz 近辺に特徴がある。なお走行音は機材の関係上 11,000 Hz までしか再現できていない。しかし雑音が聴こえる状態ですべての実験行うため問題はないと考えられる。

雑音の合成方法としては，双方のビート位置がつねに一致するのが望ましいが，実際には若干の揺らぎが存在する。そこで，20 回分のビート位置誤差の平均が 0 秒（標準偏差 0.015）になるよう合成した。

以上のように生成した実験用音声データは，実際の音楽聴取形態を想定し，ヘッドフォンで両耳

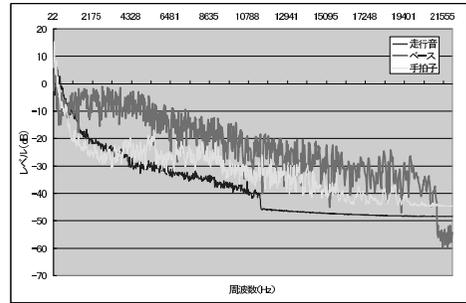


図 7 走行音とビート音の周波数特性

Fig. 7 Frequency characteristics of noise and music-beats.

へ提示する形式をとった。

**実験 1** 提案手法が何もしない場合と比べて感覚的に雑音を低減できるかを確認する実験を行った。雑音と音楽のビートを何の操作もなくバラバラに合成した音声 A，音楽の BPM を調節してビートを同期させた音声 B を用意した。この 2 つを試聴してもらい，5 段階（1：非常に気になる，5：まったく気にならない）で評価してもらった。このとき，被験者には 2 つの音が混ざっていることだけを伝え，どちらが A か B かは伏せておいた。雑音の BPM は 65 である（図 8）。

**実験 2** 提案手法の耐雑音性を評価する実験を行った。実験 1 で用いた音声 A・B 双方で混ぜ合わせる雑音の音量レベルを変え，どの時点から雑音が気になりだしたかを答えてもらった。雑音のレベルとしては，録音状態から -15，-10，-7，-5，0 dB と 5 段階に減衰させたものを用意し，雑音が気になり始める限界点を答えてもらった（図 9）。

**実験 3** 3 章で述べた手法は，雑音と音楽の BPM に大きな差がある場合は音楽の雰囲気が変わってしまうため適用できない。適用可能な曲が雑音の BPM に依存することになる。そこでビートを一致させる間隔を変更して実験を行った。音楽のビートに対して強拍に 1 つずつ合わせた音声 C，1 つとばして合わせた音声 D について，3 段階（C の方が良い，どちらも変わらない，D の方が良い）で評価してもらった。このとき，雑音の BPM は変更せず，C と D で別の音楽（C: BPM=65，D: BPM=130）を使用した（図 10）。なお，ビート楽器の周波数特性は実験 1 と同一である。

**実験 4** リズム楽器のアクセントは，ジャズやレゲエ

たとえば雑音の BPM が 120，音楽の BPM が 60 の場合，音楽の速度は 2 倍になってしまう。

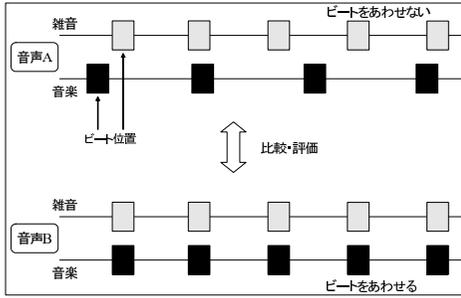


図 8 実験 1 (有効性評価)  
Fig. 8 Test 1 (availability).

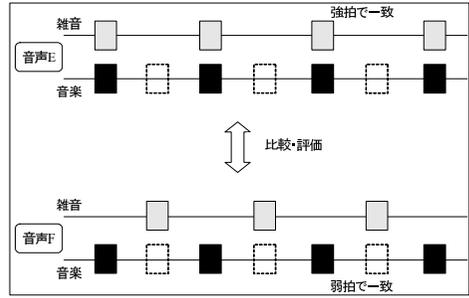


図 11 実験 4 (裏拍評価)  
Fig. 11 Test 4 (back-beat).

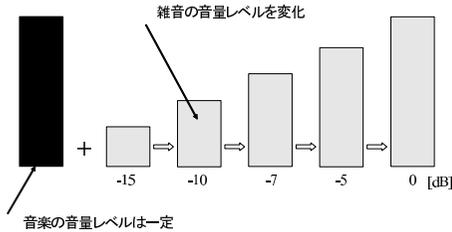


図 9 実験 2 (耐雑音性評価)  
Fig. 9 Test 2 (noise resistance).

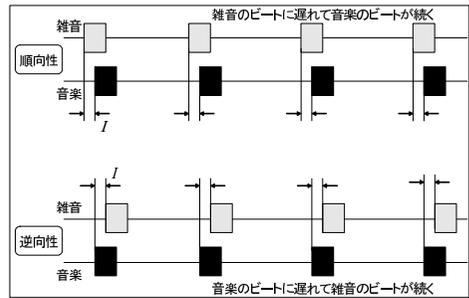


図 12 実験 5 (継時特性評価)  
Fig. 12 Test 5 (temporal characteristic).

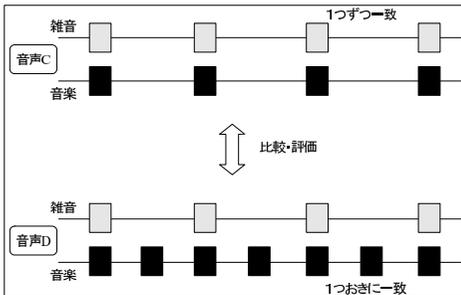


図 10 実験 3 (BPM 差が大きい場合)  
Fig. 10 Test 3 (The case where BPM difference between music and noise is larger).

などは弱拍に置くことも多い．そこでビート同期の位相を変更して実験を行った．強拍にビートを合わせた音声 E，弱拍にビートを合わせた音声 F について，実験 3 と同様 3 段階（E の方が良い，どちらも変わらない，F の方が良い）で評価してもらった．音楽は E・F どちらも同じ邦楽のポップス BPM: 65 の曲を使用した（図 11）．

実験 5 実験 1~4 はすべて，雑音と音楽ビートが同時に鳴る同時マスクング（ビート位置誤差平均が 0）に準じたものであった．そこで雑音と音楽のビート位置を意図的にずらし，ビートの提示時間特性が聴取感覚に与える影響を調べる継時マスクングに準じた実験を行った．雑音のビートに続いて音楽のビートが鳴る場合を順向性，その反対を

逆向性とし，両者の提示間隔  $I$  を変えながら，どの段階で雑音が気になりだすかの限界点を答えてもらった（図 12）．また，全体を通じた感想を自由に記入してもらった．

純音の継時マスクング実験では提示間隔  $I$  を数ミリ秒刻みで行うが<sup>9)</sup>，本実験で用いる雑音と音楽のビート位置はそれ以上の揺らぎが存在する．そこでビート位置誤差の標準偏差から， $I$  を 25，50，75，100，125 ミリ秒とした．この幅は「聞こえる/聞こえない」というマスクング量の評価には不十分であるが、「どのような音が鳴っているか」を判断する認知マスクング<sup>10)</sup>の評価には十分であると考えられる．

5.2 結果および考察

図 13 に実験 1 の評価結果を示す．横軸が評価値，縦軸がその評価値に投票した人数を表している．ビートを合わせずに合成した音声 A は「非常に気になる」（評価値 1）が多く，BPM を調節して合成した音声 B は逆に「気にならない」（評価値 4）と高い評価が多かった．それぞれの平均は A: 1.60, B: 3.55 であり，B のほうが雑音が気にならないという結果になった．

認知マスクングの影響は順向性で 40 ミリ秒，逆向性で 300 ミリ秒付近まで発生する．

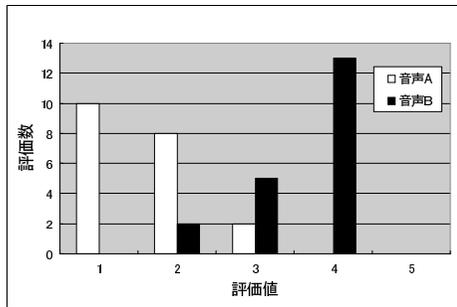


図 13 実験 1 の結果

Fig. 13 The result of test 1.

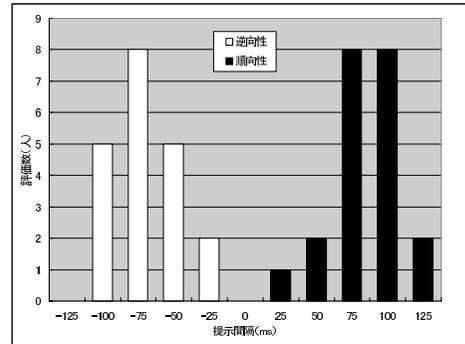


図 15 実験 5 の結果

Fig. 15 The result of test 5.

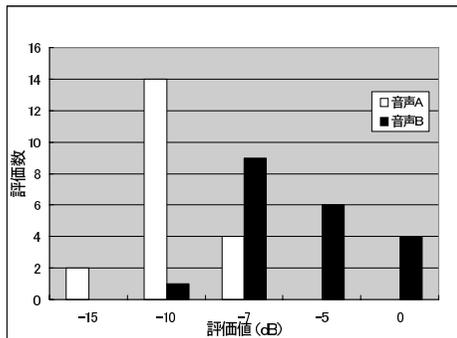


図 14 実験 2 の結果

Fig. 14 The result of test 2.

以上より、提案手法が感覚的に雑音を低減できていることが確認できた。

図 14 に実験 2 の評価結果を示す。横軸が雑音の減衰量、縦軸はその雑音が気になると評価した人数を表している。雑音の音量レベルを上げて視聴してもらったところ、音声 A と音声 B では、A: -10 dB, B: -7 dB と、A の方が早い段階で気になり出すという結果が得られた。この結果より、何の操作も加えない場合よりも大きな雑音に耐えられることが確認できた。再生音量を必要以上に大きく設定することに歯止めになると考えられる。

実験 3 の結果「C の方が良い」6 人、「どちらも変わらない」5 人、「D の方が良い」9 人となり、評価は分かれた。若干 D が高くなったのは、音楽の内容が影響していると考えられる。C は演奏楽器が少なく、ポータルとリズム楽器のみのゆったりした曲であった。これに対し D は演奏楽器が多く、つねになんらかの楽器の音が鳴っているスピーディな曲であったため、雑音がより目立たなかったと考えられる。この結果より、雑音の BPM と大きな差がある場合でも無理に変更する必要はなく、整数倍の位置で一致さえしていれば一

定の効果が得られることが示唆された。

実験 4 の結果「E の方が良い」7 人、「どちらも変わらない」8 人、「F の方が良い」5 人となった。実験 3 と同様に評価が散らばったが、変わらない、強拍、のほうが若干高くなった。これは原曲が強拍にアクセントを置いているからと考えられる。この結果より、ビートはどちらに合わせても一定の効果は得られる、ということが分かった。拍節・ビート認識技術により表裏どちらにアクセントがあるかを判定するか、ID3 タグのようなメタデータから曲のジャンルを取得することで、より効果の高いほうへ一致させると良いと考えられる。

図 15 に実験 5 の評価結果を示す。横軸が提示間隔  $I$  であり、負が逆向性、正が順向性を表している。縦軸はその雑音が気になると評価した人数を表している。それぞれの平均値は順向性が 88.75、逆向性が 70.0 となった。この結果より、逆向性のほうが聴取感覚への影響を早い段階で受けることが分かった。雑音と音楽のビートは極力一致させたほうが良いが、若干であれば音楽のビートが先行する形になっても聴取に問題はないと考えられる。

### 5.3 従来の知見との比較

次に、一般的にマスキングに関して得られている知見と比較する。岩澤によれば、継時マスキングによる認知妨害は、テスト音と雑音が同種の場合にのみ生じるとされている<sup>(11),(12)</sup>。図 7 に示すように、本実験ではテスト音であるビート楽器と、雑音である走行音の周波数特性が一致するわけではない。音楽と電車の組合せによっては、この差がさらに大きくなる可能性もある。つまり、雑音の存在は、提案手法適用にかかわらず認知可能であるといえる。

また、ドラムスやベースなどのビート音は、ミキシングにおいて中央に定位されるため、左右のチャンネル

で同位相で提供される。つまり、テスト音に位相差がある場合に知覚閾値が下がる両耳マスキングレベル差 MLD<sup>13)</sup> は発生しない。本研究はヘッドフォンで両耳提示を行っており、この点でもビート音に対する雑音の影響は大きいといえる。

それにもかかわらず、実験 1 から 3 で提案手法に一定の効果を見ることができた。これは、2.2 節および 3.1 節で述べたビート一体化による心地良さと、音楽としての意味系列の維持が、聴取感覚に好影響を与えていることを示している。

また、継時マスキング評価では、可聴閾、認知閾ともに逆向性のほうが順向性よりも雑音の影響を受けやすいことが知られている。本実験 5 でも同様の結果となった。

この理由として、ビートの遅れが音楽の系列へ与える影響があると考えられる。楽器演奏用語では、リズムが遅めになることを「モタる」と呼び敬遠している。特定のビートに対して若干遅れる程度であれば重厚なノリとなるが、本実験のようにすべてのビートに遅れるとタイミングを合わせにくくなるからである。実際、自由記入欄には「もたもたしている」「拍より長くて気持ち悪い」などの意見が寄せられていた。

順向性への影響が小さかった理由としては、走行音の波形の影響があると考えられる。今回用いた走行音は、1 つのビート「ガタン」が 2 つのピーク「ガ」と「タン」で構成されている。順向性では最初のピークが音楽のビートとずれたとしても、すぐに次のピークが一致する。また、リズムが速めになることを「走る・突っ込む」と呼び、軽快なノリにつながる。この軽快さが実験に用いた曲とマッチした可能性もある。

#### 5.4 音楽特有の効果

次に、音楽特有の現象について考察する。一般的な継時マスキング評価では、テスト音と雑音ともに 1 パルスしか用いない。これに対し本研究が想定する環境では音楽と雑音のビートが周期的に現れる。よって間隔  $I$  を大きくしていくと、雑音のピークは次第に次のビート位置に近づいていく。

この周期性の存在により、間隔  $I$  が一定以上になると、逆に聴取感覚が改善されると考えられる。実験 4 は裏拍の位置で聴取感覚が改善されることを示しているが、これは間隔  $I$  を雑音のビート周期  $T$  の  $1/2$  の地点まで広げたものといえ換えることができる。

## 6. 議 論

本章では、提案手法を実際の携帯型プレイヤー上に実現するうえでの様々な要件について議論する。

### 6.1 ハードウェア

提案手法では雑音の入力が必要になるため、プレイヤーにマイクを取り付ける必要が出てくる。左右の耳に対する到達時間差まで考慮するならば、ヘッドフォン部分の左右に取り付ける必要がある。しかし、各チャンネルを独立に処理してしまうと左右で位相のズレが生じ、ユーザにとって大きな違和感となってしまおうと考えられる。よって左右どちらか一方、もしくは体の中心線上のどこかに設置するのが良いと考えられる。

### 6.2 実際の動作

本論文では等速度運行中を仮定した実験を行ったが、実際には発車時と停車時には速度変化がともなう。走行音の BPM 変化に追従して音楽の速度を頻繁に変更してしまうと、ユーザには不安定な演奏として感じられ、逆に違和感となる可能性がある。よって BPM が安定したタイミングで提案手法を適用するのが望ましい。

また、BPM が一定になった後では、音楽の再生速度と再生位置の変更がともなう。BPM 安定後に雑音ビートの位相が変化した場合も同様である。実験 3 の結果より、ビートが整数倍の位置で一致すれば効果があるため、極端に速度を変更する必要はない。変更率が少ない速度への変換を行えばよい。実験 4 の結果より、ビートを合わせる位置は強拍・弱拍を問わず一定の効果があるため、どちらか近い方へ合わせればよい。以上より、ビートの解釈を拡大することで、速度差が 10% 以上ある場合にも提案手法を適用可能である。

以上の実現性確認のため、関東地方のある路線 12 区間で、5 秒以上走行音のビートが一定である時間を秒単位で測定した。その結果、ビートが一定だった箇所は 21 カ所、持続時間の平均は 22.4 秒（標準偏差 10.2、最小値 5 秒、最大値 50 秒）となった。1 区間での合計が 1 分以上に及ぶ場合もあった。この長さは路線の駅間距離に依存するが、現在のポップスが 3~4 分の長さであることを考えると、平均で 1 曲中の 10~12% に提案手法を適用できることになる。曲全体とはいかないものの、ギターソロやサビなど視聴者が特に聴き入る部分はカバーできる可能性がある。また、通勤時に等速運行が多い快速や急行を利用することが多いため、平均持続時間もさらに長くなるものと考えられる。以上より、本手法は電車内での音楽聴取に一定の貢献ができるといえる。

なお、この測定において、等速走行中におけるビートのタイミング（位相）変化が 5 カ所で確認された。この原因は、鉄橋など線路の設置状況が変わる地点であった。この位相変化が短時間で頻繁に発生するなら

ば、本手法は適用できないことになる。線路の設置場所によるので一概に判断はできないが、本調査を行った路線では位相変化箇所は少なかった。さらに、先の平均持続時間は、このような位相変化の前後を別々のものとして計算したものである。位相変化の追従に数秒を要するとしても、その前後で平均 40 秒近く本手法を適用可能である。

### 6.3 既存手法との併用性

提案手法そのものは、雑音を残しつつ耳障りにならないようにするものであるが、既存の手法と併用することで雑音の消去効果を向上させることができると考えられる。

まず、密閉型ヘッドフォンの形状を変更することなく共存が可能である。また、雑音に対する処理をいっさい行わないので、ノイズキャンセリングヘッドフォンとも共存可能である。よって既存の製品で雑音を低減したうえで提案手法を適用すれば、ボリュームを上げることなくほとんど知覚できないレベルまで雑音の低減が可能だと考えられる。

### 6.4 その他の適用場面

本論文では電車の走行音を対象としたが、ビートが一定の時間は限られるうえで、新幹線などつなぎ目のないレールも存在する。そこで、その他の適用場面に ついて 3 つ例をあげる。

1 つは、高速道路における継ぎ目の走行音である。これは携帯用ではなく車載プレイヤーが対象になる。電車の走行音に比べて仮想ビートの BPM は小さめではあるが、ほぼ一定周期で耳障りな音が発生する顕著な例である。

もう 1 つは、ジョギングにおける足音や衣擦れ音である。現在、携帯型プレイヤーで音楽を聴きながらジョギングをする人は非常に多い。この例では、音楽に合わせてランナがペースを変えるのではなく、ランナのペースに合わせて音楽のテンポが自動調整される。これは聴きやすさだけでなく、ランナの動きとの一体感も向上するため、ジョギング時の快適性も向上させる可能性がある。

最後は、ユーザの近くで流れる他の音楽である。たとえば電気店などではアップテンポな販売促進用の音楽をつねに大音量で流している。また、常時ラジオや BGM を流している場所もある。このような音楽のうち、特にビートを構成する低音は、ヘッドフォンをしていても耳に入ってくることが多い。

## 7. おわりに

本研究では、楽器演奏の知見を用いて雑音を音楽に

溶け込ませ、その影響を感覚的に低減する手法を提案した。評価実験では、ビート位置を一致させたほうが雑音が気にならないこと、ビート位置を厳密に一致させなくても一定の効果が得られることを確認した。この結果より、雑音の感覚的低減に効果があるという有効性と、極端に音楽の再生速度を変更しなくてもよいという柔軟性を示した。

一方で、本論文では理論の検証を主題においたため、現段階では外部雑音の周期性を自動認識するには至らなかった。認識を困難にしているのは、車両・線路間の個体差と、慣性や風による車体の振動音である。個体差に関しては、より多くの走行音サンプルを収集したうえで、文献 4)、5)、14) など周波数変化をとらえる手法が適用可能ではないかと考えている。その他の振動音に関しては、同様にサンプルの収集と分析により、その特徴を明らかにする必要がある。音声処理だけで難しい場合は加速度計を併用し、水平方向の揺れによる騒音の影響を小さくするなど考えられる。

また、ビートは音響的信号以外にも存在する。1 つは生体的なビートである。これは心拍であったり、歩クリズムであったりする。もう 1 つは視覚的なビートである。走行中の電車において、窓の外を流れていく電柱の様子などがこれにあたる。このようなビートは音楽的ビートの知覚に影響を及ぼすことも指摘されている<sup>15)</sup>。

これら様々なビートを同期させることができれば、屋外で快適な視聴を約束するプレイヤーが実現するに違いない。従来はヘッドフォンなど音声の出力デバイスに重きを置いていたが、今後は音響的ビートを認識するマイクロホン、視覚的ビートを認識するカメラ、生体的ビートを認識するセンサなど、プレイヤーに入力デバイスを接続することが重要になるであろう。

謝辞 本研究の一部は平成 16 年度選定文部科学省私立大学学術高度推進事業（オープン・リサーチ・センター整備事業）の支援のもとに行われた。

## 参考文献

- 1) Schafer, R.M.: 音さがしの本—リトル・サウンド・エデュケーション, 春秋社 (1996).
- 2) 岩宮眞一郎: 音楽と映像のマルチモーダル・コミュニケーション, 九州大学出版会 (2000).
- 3) 井上亮文, 天野玲奈, 備瀬翔平, 市村 哲, 松下 温: 携帯型音楽プレイヤーのためのコンテンツ再生・配信方式, 情報処理学会研究報告 2006-GN-058, pp.133-138 (2006).
- 4) 吉井和佳, 後藤真孝, 奥乃 博: テンプレート適応を利用した実世界の音楽音響信号に対するド

ラムスの音源同定, 情報処理学会研究報告 2003-MUS-53, pp.55-60 (2003).

- 5) 吉井和佳, 後藤真孝, 奥野 博: 単一テンプレート適応法による音楽音響信号を対象としたハイハットシンバルの音源同定, 情報処理学会研究報告 2004-MUS-56, pp.49-56 (2004).
- 6) Roucos, S. and Wilgus, A.M.: High Quality Time-Sceme Modification for Speech, *IEEE International Conference on ICASSP'85*, Vol.10, pp.493-496 (1985).
- 7) Wong, J.W., Au, O.C. and Wong, P.H.: Fast Time Scale Modification using Envelope-Matching Technique (EM-TSM), *IEEE International Symposium on ISCAS'98*, Vol.5, pp.550-553 (1998).
- 8) 小川 要: ソフトウェアによるオーディオデジタル信号処理のレシピ, *C MAGAZINE*, Vol.3, pp.13-41 (2002).
- 9) 城戸健一 (編): 音響工学講座 1 基礎音響工学, コロナ社 (1990).
- 10) Massaro, D.: Preperceptual images, processing time and perceptual units in auditory perception, *Psychological Review*, Vol.79, No.2, pp.124-145 (1972).
- 11) 岩澤秀紀: 逆向認知マスクングと聴覚の多チャンネル処理, 日本音響学会聴覚・音楽音響研究会資料, pp.H-93-78, MA93-31 (1993).
- 12) 岩澤秀紀: 聴覚逆向認知マスクングにおけるマスクの有効周波数範囲, 日本音響学会平成7年春季研究発表会後援論文集, pp.447-448 (1995).
- 13) Moore, B.: 聴覚心理学概論, 誠信書房 (1994).
- 14) 後藤真孝, 村岡洋一: 打楽器音を対象にした音源分離システム, 電子情報通信学会論文誌, Vol.J77-D-II, No.5, pp.901-911 (1994).
- 15) 長嶋洋一: 音楽的ビートが映像的ビートの知覚に及ぼす引き込み効果, 芸術科学会論文誌, Vol.3, No.1, pp.108-148 (2004).

(平成 18 年 6 月 21 日受付)

(平成 18 年 12 月 7 日採録)



井上 亮文 (正会員)

1999年慶應義塾大学理工学部計測工学科卒業。2001年同大学院前期博士課程修了。2005年同大学院後期博士課程修了。博士(工学)。現在、東京工科大学コンピュータサイエンス学部助手。コンテンツ記録・再生, ヒューマンインタフェース, ネットワークセキュリティの研究に従事。ヒューマンインタフェース学会, ACM 各会員。



備瀬 翔平

2006年東京工科大学工学部情報通信工学科卒業。在学中, 音楽情報処理の研究に従事。



市村 哲 (正会員)

1966年生。1989年慶應義塾大学理工学部計測工学科卒業。1994年同大学院理工学研究科博士後期課程修了。博士(工学)。同年富士ゼロックス(株)入社。1997~1999年富士ゼロックスパロアルト研究所駐在。2002年4月より東京工科大学助教授。グループウェア, ネットワークサービス, 生体情報活用等の研究に従事。『IT TEXT 基礎 Web 技術』, 『IT TEXT 応用 Web 技術』(オーム社)。DICOMO 2003, 2005, 2006 優秀論文賞受賞。ACM, 電子情報通信学会各会員。



松下 温 (フェロー)

1963 年慶應義塾大学工学部電気  
工学科卒業．1968 年イリノイ大学  
大学院コンピュータサイエンス専攻  
修了．工学博士．1989 年より 2002  
年 3 月まで慶應義塾大学理工学部教

授，2002 年 4 月より東京工科大学教授および慶應義  
塾大学理工学部客員教授，2003 年 4 月より 2006 年  
3 月まで東京工科大学コンピュータサイエンス学部学  
部長および慶應義塾大学理工学部客員教授，2006 年  
4 月より住宅情報化推進協議会．マルチメディア通信，  
コンピュータネットワーク，グループウェア等の研究  
に従事．情報処理学会理事，同学会副会長，マルチメ  
ディア通信と分散処理研究会委員長，グループウェア  
研究会委員長，電子情報通信学会情報ネットワーク研  
究会委員長，MIS 研究会委員長，バーチャルリアリテ  
ィ学会サイバースペースと仮想都市研究会委員長，情  
報処理学会 ITS 研究会委員長等を歴任．郵政省，通産  
省，建設省，農水省，都市基盤整備公団，行政情報シス  
テム研究所等の委員長，座長，委員等を多数歴任．特  
に国土交通省，住宅情報化標準策定委員会委員長，経  
済産業省総合エネルギー調査会電子計算機と磁気ディ  
スク委員会委員長を務める．現在，経済産業省総合エ  
ネルギー調査会ルータ装置基準委員会委員長，最高裁  
判所専門委員．『やさしい LAN の知識』（オーム社），  
『201x 年の世界』（共立出版）等著書多数．1993 年情  
報処理学会ベストオーサ賞，1995 年および 2000 年情  
報処理学会論文賞，2000 年 10 月 20 日情報処理学会  
40 周年記念 90 年代学会誌論文賞，2000 年 10 月 2 日  
電子情報通信学会フェロー，2000 年 10 月 VR 学会サイ  
バースペース研究賞，2001 年 5 月情報処理学会功  
績賞，2002 年 3 月情報処理学会フェロー，電情報通  
信学会，人工知能学会，ファジィ学会，IEEE，ACM  
各会員．

---