

音漏れモデルに基づくスマートフォンを用いた 音漏れ検出手法

中 久治¹ 土井 敦士¹ 根岸 佑也² 梶 克彦³

概要：スマートフォンの普及により電車内やカフェなどの公共空間で音楽や動画、ゲームを手軽に楽しめるようになった。それに伴い、イヤホンからの音漏れのような公共空間ならではの迷惑行為も、より顕著な問題になりつつある。音漏れは周囲の環境音より大きいイヤホンからの漏出音であると捉えられ、音楽を楽しむ本人が音漏れを聴けず、自身で音漏れに気がつきにくい。本研究では、公共マナーの向上をねらい、スマートフォン自体が再生している音に関して音漏れ検出できる機能の実現を目的とする。しかしながら、音漏れしやすい音域は、イヤホンと環境音に依存して変化する。また、再生する楽曲によっても変化する。そこで本稿では、再生楽曲の周波数特性のうち、イヤホンと環境音ごとの音漏れ特性を考慮し、音漏れする可能性が高い周波数のみに注目して、マイクからの観測音と比較することで、効果的に音漏れを検出する手法を提案する。実際に、スマートフォンを用いた楽曲の再生と録音を行い、イヤホンや環境音毎の音漏れ特性モデルを構築し、提案手法を用いた実験を行い、音漏れ検出が可能であることを確認した。

A Sound Leakage Detection Method for Smartphone Based on Otomore-Model

Hisaharu Naka¹ Atsushi Doi¹ Yuya Negishi² Katsuhiko Kajii³

1. はじめに

迷惑な社会問題の1つとして、音漏れがある。スマートフォンやポータブルオーディオプレーヤの普及により、電車内やカフェなどの公共空間で音楽や動画、ゲームを楽しむようになった。しかし、イヤホン等からの音漏れが人を不快にさせる事例も増えた。音楽を楽しんでいる本人が自身で音漏れを聴けないため、音漏れが発生しているかの判断は一人では困難である。もし、スマートフォン自体に再生している音に関して音漏れ検出できる機能が備われば、公共マナーの向上につながる可能性が高い。本研究では、そのようなスマートフォンを用いた音漏れ検出機能の実現を目的とする。

本稿において、他者に迷惑となる音漏れを周囲の環境音より大きいイヤホンからの漏出音と定義する。周囲の環境音より小さいイヤホンからの漏出音は、環境音に混ざり迷

惑になりづらい。例えば、電車などのうるさい環境で気にならなかった音漏れが、話し声や機械の動作音などの少ない比較的静かな環境(以降静音環境と呼ぶ)では気になる。したがって、イヤホンの種類によって異なる出力周波数特性や周囲の環境音の周波数特性によって、迷惑になる音漏れは変わると考えられる。

既存の音漏れの検出手法として、流れる音楽のリズムに着目し音漏れを検出する手法 [2] などがあるが、着目するリズムが環境音に混ざりづらい周波数帯域であるか、イヤホンの出力特性を考慮すると、より効果的な音漏れ検出の実現を期待できる。

本稿では、イヤホンごとの音漏れ特性の違いを考慮した音漏れの検出手法を提案する。提案手法の特徴は次の通りである。まず、音漏れの検出の前段階として、ユーザのイヤホンと周囲の環境音における音漏れしやすい周波数帯域を、スマートフォンのマイクを用いてイヤホンごと環境ごとに解析した音漏れ特性モデルを作成する [1]。音漏れ特性モデルでは、GMM [7] を用いることにより、周波数帯域毎

¹ 愛知工業大学大学院経営情報科学研究科

² メタプロトコル株式会社

³ 愛知工業大学情報科学部

の音漏れしやすい度合いを測定ノイズなどを除去し、軽量のパラメータで特性を近似できる。続いて、音漏れ検出時には、再生楽曲が周囲に音漏れすると仮定する場合に、音漏れ特性モデルを用いて環境音に混ざりづらく、かつ、使用イヤホンで音漏れしやすい周波数帯域を数箇所求めておく。マイクからの観測音の音漏れの周波数特性と、その周波数帯域のみの類似度を比較することで、ユーザの状況下に応じた、効果的な音漏れ検出を可能にする。再生する楽曲に応じた、音漏れの仮定と観測値の類似度の値は一定ではないため、最終的にユーザに音漏れの有無を提示する際には、再生楽曲毎に類似度の下限と上限を見積もり、音漏れレベルを正規化する。実際に、環境とイヤホンが異なる場合の音漏れ特性モデルの構築と、音漏れ検出手法の実験を行い、提案手法の有効性を確認した。

本稿の構成は以下の通りである。2章で音漏れの定義を行い、本稿の目標とする音漏れの検出手法について述べ、課題点を述べる。3章で既存の音漏れに関する研究について述べる。4章で音漏れ特性をモデル化する手法について述べる。5章で提案する音漏れの検出手法の詳細な手順について述べる。6章で提案手法の有効性を検証する実験について述べ、7章でまとめと今後の課題を述べる。

2. 音漏れ検出のための課題整理

2.1 音漏れの定義

本研究にて検出対象にする他者に迷惑な音漏れについて、具体例を示しながら、どのようにしてイヤホンから漏出するかを説明する。図1に、電車内で27.5Hzから21705Hzまでの各周波数のsin波をスマートフォンに接続したイヤホンより流し、スマートフォンのマイクで録音をした結果と、電車内の環境音の周波数特性を示す。再生時の音量は、低周波数帯域の音が聞こえる程度とした。図1の緑枠内に注目すると、高周波数帯域の音量がsin波を流した方が環境音に比べ大きい。この理由は、イヤホンから漏れ出る音が周囲の環境音より大きくなり、音漏れが発生しているためと考えられる。もし、音楽を電車内の環境音に負けないように音量を上げると、イヤホンから漏れ出る高周波数帯の音が環境音より大きくなり、不快なシャカシャカした音漏れ(図1緑枠内)が目立つと考えられる。

以上の周囲の環境音より大きいイヤホンからの漏出音を音漏れと定義する。したがって、環境音によって音漏れの内容は変化する。電車などのうるさい環境で環境音によってマスキングされて気にならなかった音漏れが、静音環境ではマスキングされずに音漏れとして他者が気になる可能性が高い。逆に、環境音にマスキングされる小さい音漏れは、他者に迷惑にならず、音漏れ検出の対象にしなくても問題ない。

また、イヤホンの種類によって出力する周波数特性が異なる。加えて、イヤホンの開放・密閉型など構造の違いや

カナル型やインナーイヤー型など装着方法の違いがあるため、イヤホンによって音漏れしやすい周波数帯域が異なる。

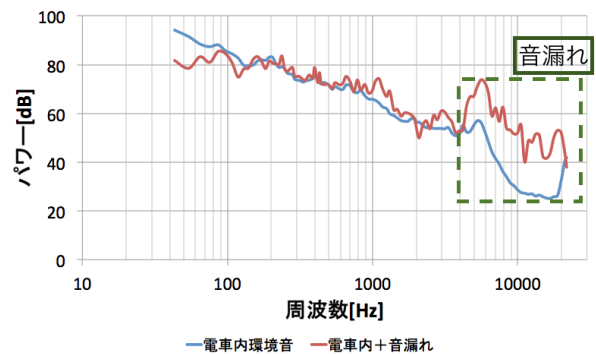


図1 電車内の環境音とイヤホンからの音漏れの比較

2.2 音漏れ検出可能なスマートフォンアプリの検討

本節では、ユーザが自身のスマートフォンを用いて、スマートフォン自体で再生している音について音漏れ検出可能なアプリケーションについて検討する。図2に、本研究で目標とする音漏れ検出可能な音楽プレーヤーアプリを示す。想定アプリでは、スマートフォンのマイクで周囲の音を観測できるとする。これにより、イヤホンの音漏れしやすい周波数特性と周囲の環境音の周波数特性、再生する楽曲の周波数特性を用いて、音漏れすると仮定した場合の観測音を推測できる。音漏れを含むと仮定した観測音とマイクで観測した実測の音を比較することで音漏れを検出し、アプリケーション上でユーザに注意を呼びかける。

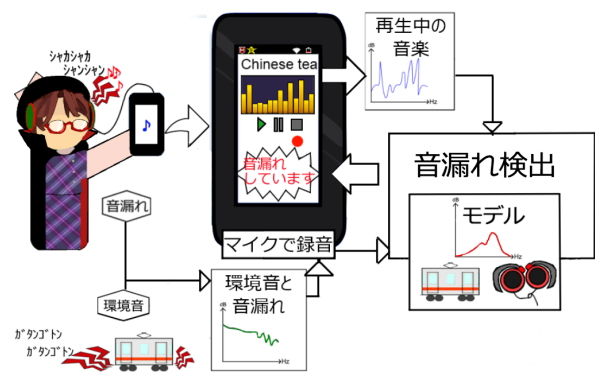


図2 音漏れ検出・抑制の模式図

2.3 課題の整理

前節までの議論をふまえ、スマートフォンを用いた音漏れ検出について、主な課題2点を以下にまとめる。

- (A) **音漏れ特性モデル**: イヤホンによって異なる音漏れのしやすさ特性のモデルと、モデル化を容易に行う手法が必要である
- (B) **音漏れ検出手法**: 再生楽曲が周囲に音漏れすると仮定

した場合に、音漏れ特性モデルを用いて環境音に混ざりづらく、かつ、使用イヤホンの音漏れしやすい周波数帯域を十分に考慮した手法が必要である

さらに、(A)の音漏れ特性モデルでは、スマートフォンのマイクを使用し簡易的に測定を行うため、測定誤差を適切に除去できるように、軽量のパラメータで近似的できるようなモデル表現方法が望ましい。軽量の表現は、(B)の音漏れ検出をリアルタイムに実現するために、計算負荷の抑制の面からも有効である。それぞれ、課題(A)に対応する手法を4章にて述べ、課題(B)については5章にて述べる。

3. 関連研究

既存の音漏れの検出には、再生コンテンツのリズムに着目し、ラベリアマイクを用いて音漏れを検出する手法が提案されている [2]。ラベリアマイクはマイク付きのイヤホンである。この手法では、音漏れとして気になる音がリズムを有する音と捉えて検出をする。もし、前章にて述べたように音漏れしにくい周波数帯域のリズムに着目してしまうと、音漏れを検出しにくい可能性が高い。

また、音漏れの抑制手法として信号の最大値を制御するAVLS(Automatic Volume Limiter System)や、信号内に含まれる楽器群の音量と定位を操作し、漏出しやすい帯域を低減し音漏れを減らす手法がある [3]。あるいは、アクティブノイズコントロールを使う手法 [4] や、他の音で不快な音を打ち消す手法 [5][6] がある。しかし、いずれの手法も周囲の環境やイヤホンの違いなど、検討手法における課題点を考慮していない。

4. イヤホンの音漏れ特性のモデル化

4.1 音漏れモデル

本章では2.3節の課題(A)について、イヤホンによってどの周波数帯が音漏れしやすいのか、音漏れ特性を表現する方法について説明する。周波数毎に音漏れのしやすさを順番に調べる場合、図3のように、実測には環境音や測定誤差が含まれ、滑らかな周波数スペクトルを得られない。また、測定した周波数の間の補間も行いたい。したがって、パラメータによって簡易なモデルとして近似できると望ましい。本稿では、周波数毎の音漏れのしやすさをGMMによって近似表現する。周波数を指数に、音漏れのしやすさを参照する際にも少ない計算量で済む。

4.2 モデル化手法

使用するイヤホンにおいて、各周波数毎に音漏れしやすさを測定し、GMMで近似表現した音漏れ特性をモデル化する手法を以下に示す。本手法では、スマートフォンを用いて各周波数のsin波を逐次切り替えながらマイクで測定していきモデル化する。その際に、環境音を含めて録音し

てしまうため、環境音をキャンセルするための準備(以下のStep.1)と、EMアルゴリズムによるGMMへの近似から構成する。また、周波数をそのまま線形的にGMM中の変数とせず、平均律に基づき27.5Hzを基準とした対数にて換算する。

$$f = \log_2\left(\frac{f_i}{27.5}\right) * 12 \quad (1)$$

Step.1 音を再生せずに環境音のみを録音し、FFTより周囲の環境音の周波数スペクトルとする。

Step.2 再生するsin波の周波数を変化させながら、環境音を含めた音漏れの音量を一定時間測定し、各周波数の測定結果を結合して周波数スペクトルを得る。(例:図3赤線)

Step.3 前手順の周波数スペクトルより、環境音の周波数スペクトル分を差し引く。(例:図3青線) 値が負となった箇所はその音は微小であるとみなし、値を0とする。差し引き後の周波数スペクトルが使用イヤホンにおける音漏れ特性に相当する。(例:図4)

Step.4 EMアルゴリズム [7] を用いてGMMに近似する。EMアルゴリズムに与える初期のGMM中のガウス分布のパラメータ(平均と分散)と、混合数は、事前にX-means法 [8] により決定する。GMM化した結果、音漏れしやすい周波数特性の周波数がGMMの分布関数の変数に対応し、分布関数の値が各周波数の音漏れしやすい度合いに対応する。

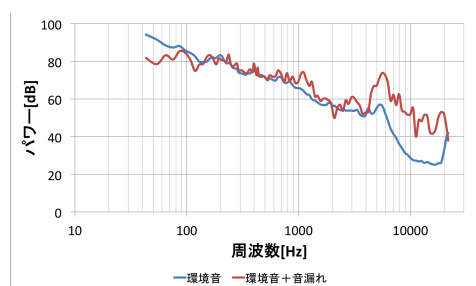


図3 環境音と音漏れとの比較

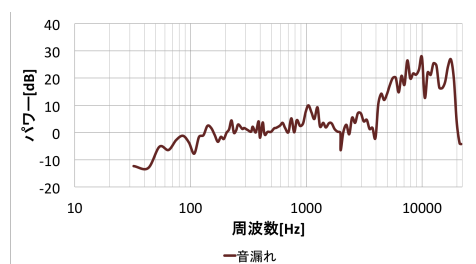


図4 音漏れの周波数スペクトル

4.3 実験 手順

実際に、一般的な室内のような静音環境と電車内の環境において、GMMによる音漏れ特性のモデルを作成する実験を行った。実験に用いた機器を表1に示す。各環境において、イヤホンを装着した位置から20cm離れた場所にスマートフォンのマイクを配置(図5)する。録音内容として、音量が一定で、27.5Hzから21096.2Hzまで1秒ごとに平均律に基づき半音ずつ上がるsin波を102種類再生した。再生音量は、イヤホンやヘッドホンが人の耳に当たる箇所に騒音計を近づけ、4186Hzのsin波が78dBを示すように揃えた。FFTのサイズは4096点にした。

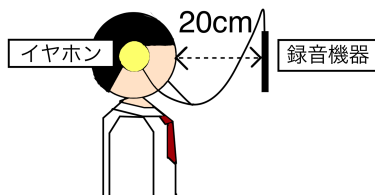


図5 音漏れ特性モデル化の実験構成

表1 実験に用いた機器

録音機器	AQUOS PHONE(SBM203SH)
イヤホン	EQ500(audio-technica)-耳掛け型 900ST(SONY)-オーバーヘッド型

結果：環境音やイヤホンによって異なる音漏れ特性

実験の結果として、静音環境で耳掛け型の音漏れ特性モデルを図6に、電車環境で耳掛け型の音漏れ特性を図7、静音環境でオーバーヘッド型の音漏れ特性を図8に示す。GMMによる近似で、測定誤差による細かな変動を除去できている。

各図中の音漏れ特性を比較すると、周囲の環境音の違いとイヤホンの違いによって音漏れしやすい周波数帯域が異なる様子を確認できる。例えば、静音環境(図6)に比べて電車内(図7)では音漏れの音量が全体的に小さく、かつ、電車の音にマスキングされやすい低周波数帯では音漏れしにくいと言える。同じ静音環境下におけるイヤホンの違いとしては、オーバーヘッド型(図8)で最も音漏れしやすい周波数帯域が7000Hz付近にある一方で、耳かけ型(図6)では3000Hz付近にある。他にも、耳掛け型に比べ、オーバーヘッド型は10000Hz付近の音漏れする音量が小さい。

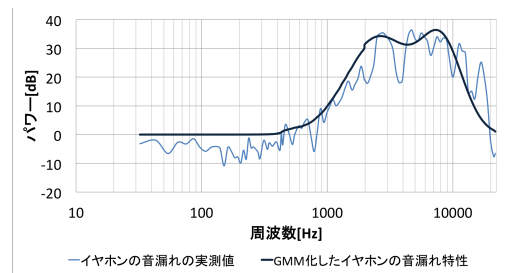


図6 静音環境で耳掛け型の音漏れの実測値とモデル

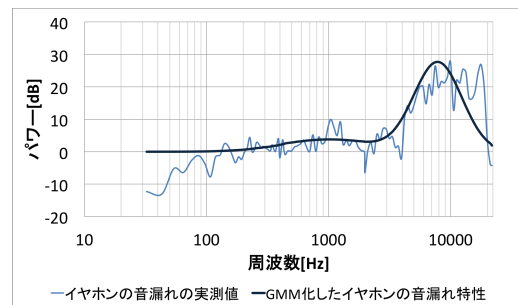


図7 電車環境で耳掛け型の音漏れの実測値とモデル

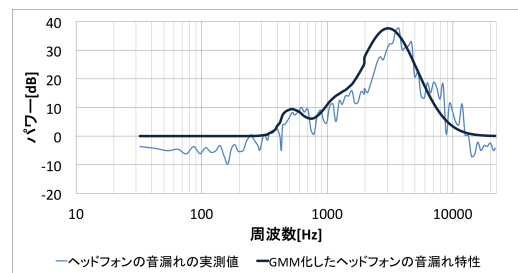


図8 静音環境でオーバーヘッド型の音漏れの実測値とモデル

5. 音漏れ検出手法

前章で述べたイヤホンの音漏れ特性モデルを応用すると、もし、再生する楽曲が音漏れすると仮定する場合には、最も音漏れとして観測しやすい周波数帯域を絞り込み、効果的な音漏れ検出を期待できる。本章では、提案する音漏れ検出手法について、手順を説明する。

検出手法の大枠の流れを図9に示す。各ステップについて詳細を述べる。

Step.0 環境音の取得

楽曲を再生する前に環境音のみを録音し、環境音の周波数特性とする。

Step.1 音漏れを仮定した周波数特性を生成

楽曲を再生開始後、一定間隔で、再生する楽曲のフレームの周波数スペクトルに音漏れ特性モデルを加味し、再生楽曲が音漏れしていると仮定する場合の観測音の周波数スペクトルを生成する。また、最大パワーが1になるように正規化する。

具体例を、図10, 11, 12に示す。再生楽曲のある瞬間の周波数スペクトル(図10)に対して、音漏れモデルの最大音量を1として正規化したモデル(図11)を掛け合わせ、

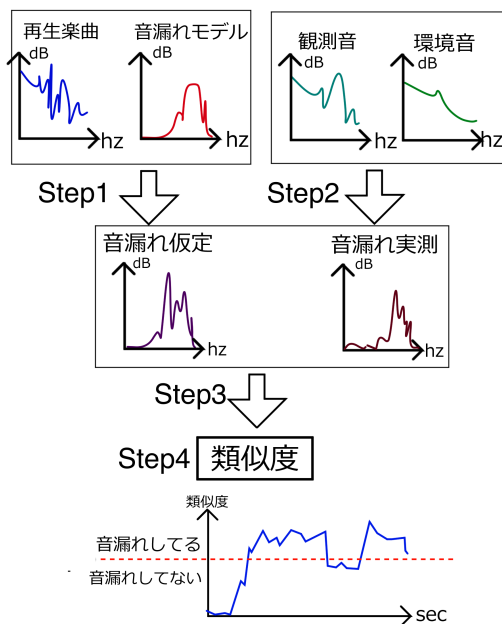


図 9 音漏れ検出のフローチャート

音漏れしたと仮定した場合の周波数スペクトル (図 12) を得る。

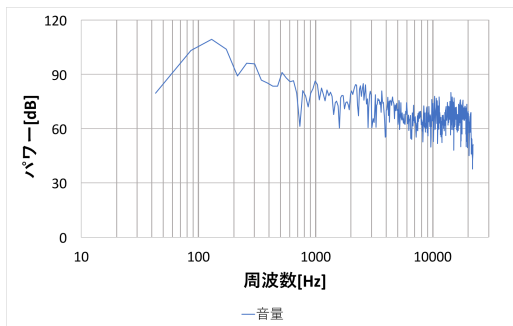


図 10 再生楽曲のとある瞬間の周波数スペクトル

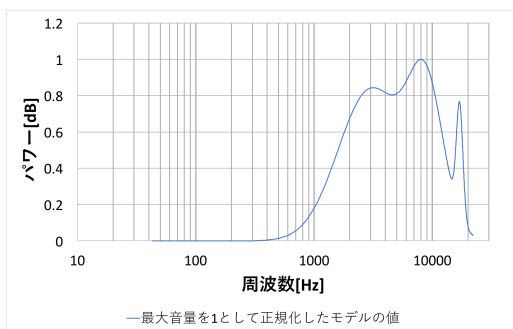


図 11 音漏れモデル (最大箇所を 1 に正規化)

Step.2 マイクによる観測音の取得

音漏れの仮定を求めたフレームを再生した時刻の音を、マイクで録音し、周波数スペクトルを得る。仮定と同様に、最大パワーが 1 になるように正規化する。

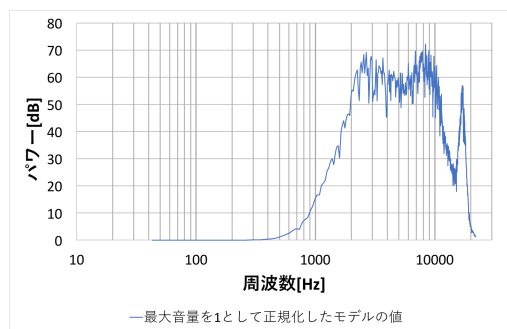


図 12 音漏れしたと仮定した場合の観測音

Step.3 音漏れ仮定と観測音の比較

再生楽曲と音漏れ特性を考慮して、音漏れしていると仮定した場合の周波数スペクトルの内、音漏れしやすい周波数として注目すべき点を n 個、選び出す。選び出す方法として、スペクトル包絡を求め、ピーク部分に対応する周波数がより高い値を持つ順に上から n 点取る。これにより、楽曲ごとの音漏れのしやすさを捉える。

選び出したピーク周波数毎に音漏れの仮定と実測音の値を比較し、それぞれの周波数毎の距離指標値の総和を求め、両者の類似度を算出する。類似度の計算式を式 2, 3 に示す。式 2 で仮定 a と実測 o の注目すべき周波数毎の二乗差を求め、逆数を式 3 で注目点の数だけ合計して類似度 S にする。類似度の範囲は、最も仮定と実測が一致する場合に n になり、最も一致しない場合に 0 に対応する。

$$D = (a - o)^2 \quad (2)$$

$$S = \sum_{i=1}^n \frac{1}{D_i + 1} \quad (3)$$

Step.4 音漏れ度合いの算出

Step.3 の類似度は、環境音や再生する楽曲によって変動するため、再生楽曲と環境音の状況下毎に事前に見積もった類似度の下限 S_{min} と上限 S_{max} より、0 から 1 の範囲に正規化する (式 (4))。範囲を超える値は、打ち切りによって収める。

$$S_N = \frac{S - S_{min}}{S_{max} - S_{min}} \quad (4)$$

なお、類似度および音漏れ度合いは細かく振動するため、平滑化フィルタに通す。

類似度の下限 S_{min} は、再生楽曲中で最大音量になっているフレームにて音漏れを仮定したものと、環境音のみを録音した音との間で求めた類似度とする。これは、観測音に全く音漏れが含まれない場合の S の値に対応する。

上限 S_{max} は、最も音漏れしている場合の S に対応するが、今回は S_{min} に音漏れ検出感度係数 s を乗じた値とした。音漏れ検出感度はユーザが好みに調整できる。

6. 評価実験

6.1 実験

音漏れ検出手法の有効性を検証するため、実際に3種類の楽曲について音漏れ検出を行った。イヤホンに表1の耳かけ型を使用し、静音環境で行った。検出に使用する音漏れモデルは同条件で作成したモデルを使用した。再生する楽曲は、曲中で激しい曲調と静かな曲調が現れる楽曲を選んだ。十分に音漏れするようにスマートフォンの音量を設定し、再生中は音量を変更しないようにした。楽曲再生前の10秒を類似度の下限を仮定するための時間とし、また、再生前の3秒間の周波数スペクトルの平均を取り環境音の周波数特性に用いた。以降の環境音の特性はこれを利用した。類似度の算出時に注目する周波数の数 n は6点にした。類似度の上限は下限の1.16倍に設定し、音漏れ度合いを算出した。

また、本手法によって、環境音と再生楽曲、イヤホンの音漏れ特性を十分に考慮できているか検証するために、各時点で音漏れ仮定と観測音を比較する際に注目した周波数の妥当性を評価した。妥当性としては、音漏れしやすさに対応する音漏れ特性モデル中の注目周波数の値の和とした。以上それぞれの結果を楽曲の再生前の時刻からグラフとして推移を確認する。

6.2 結果

まず、音漏れモデルを使った場合の音漏れ度合いの推移、音量の推移、妥当性をそれぞれ図13、図14、図15に示す。また、音漏れモデルを使わない場合の音漏れ度合いの推移、音量の推移、妥当性をそれぞれ図16、図17、図18に示す。加えて、それぞれのグラフ中にその時刻での音漏れの有無を大まかに記す。



図13 音漏れモデルを使用した音漏れ度合いの推移

音漏れ度合いと音量の関係を確認するため、図13と図14に注目する。再生中にスマートフォン上の音量を変更していないため、曲中で音量変化があれば、音漏れ度合いに

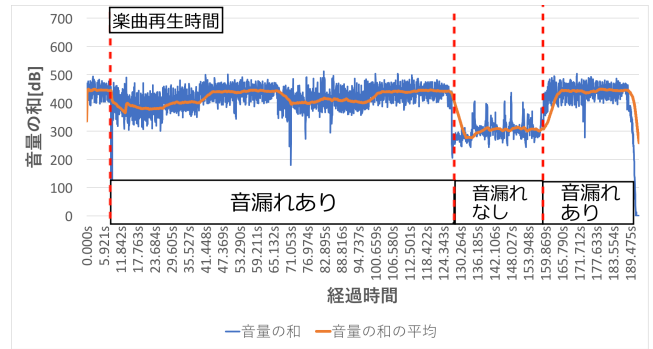


図14 音漏れモデルを使用し選んだ周波数の音量の和の推移

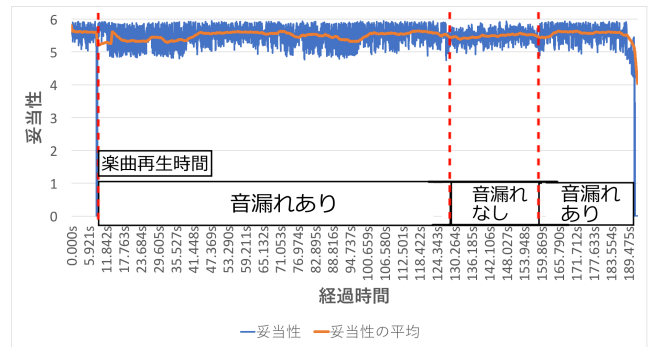


図15 音漏れモデルを使用し選んだ周波数の妥当性の推移

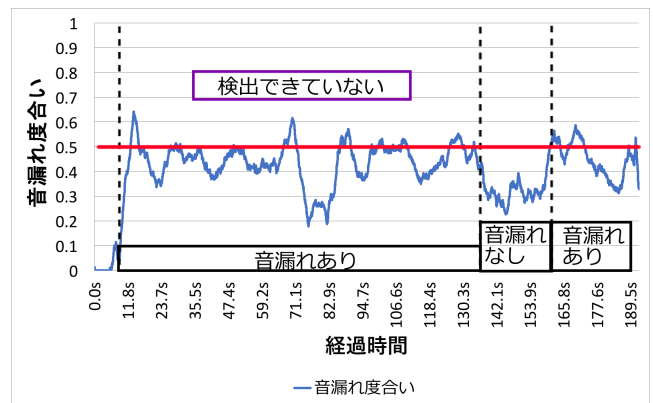


図16 音漏れモデルを使用しない音漏れ度合いの推移

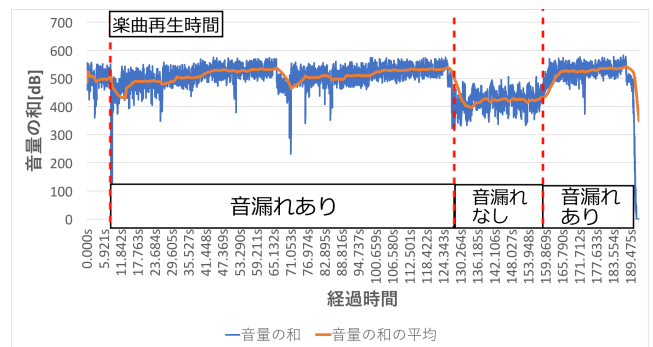


図17 音漏れモデルを使用せず選んだ周波数の音量の和の推移

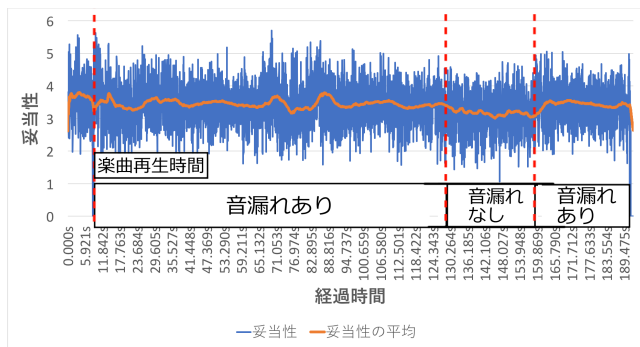


図 18 音漏れモデルを使用せず選んだ周波数の妥当性の推移

も変化があると考えられる。楽曲再生時間以降に注目すると、図 14 の音量変化に合わせて図 13 も増減しているとわかる。ここで、図 16 と図 17 に注目する。図 17 では図 14 に比べて全体的に音量は大きい、音量の増減の推移は似ている。しかし、図 16 は図 13 に比べ音漏れの有無で差異が見られない。そのため、音漏れモデルありの音漏れ度合いは音漏れモデルなしの音漏れ度合いに比べ、音漏れの状況を反映できているとわかる。

選んだ周波数が妥当かを確認するため、図 15 と図 18 に注目する。図 18 に比べ図 15 の方が、終始高い妥当性を示している。故に、音漏れモデルを用いないスペクトル包絡に比べ、音漏れモデルを用いたスペクトル包絡の方がより妥当に音漏れしやすい周波数帯域を選んでいるとわかる。

以上から、音漏れモデルを用いた音漏れ度合いは、音漏れしやすい周波数帯域を選べており、音漏れの音量に追従していると考えられる。改めて、図 13 に注目すると、図 13 では音漏れありの主観的評価に対し、類似度も 0.5 以上を示し音漏れを検出できている。また、音漏れの無い時間帯は音漏れ度合いもおおよそ 0.5 未満となりほぼ誤検出のない結果とわかる。

ここで、同様にして他の楽曲で解析した音漏れ度合いの推移を音漏れの有無の主観的評価と共に図 19 と図 20 に示す。図 19 では、一部の区間で音漏れがあるにも関わらず、音漏れを検出できていない時間(図 19 中緑枠)がある。この区間の音漏れの音量は、主観的評価で小さめであり、類似度の上限が高かったのではないかと考えられる。同様に、図 20 においても主観的評価で小さめな音漏れが全体的にあるにも関わらず、類似度は一意に 0.5 以上とならない結果を示している。この音漏れに対しても、類似度の上限は高いと考えられる。類似度の上限を下げると、音漏れ度合いは必然的に上昇するため、以上の結果から、適切な類似度の上限設定を行うと音漏れの検出が可能であると考えられる。

7. まとめ

本稿では、環境やイヤホンの違いを考慮した音漏れの検

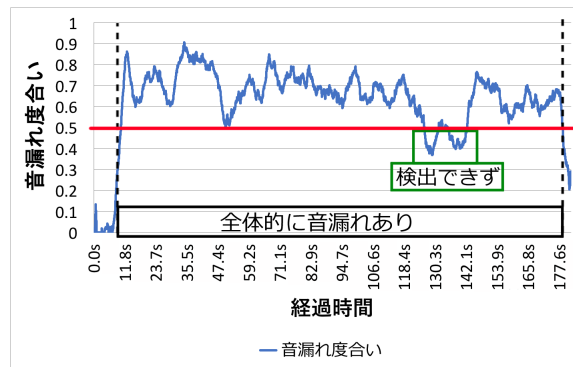


図 19 音漏れモデルを使用した音漏れ度合いの推移 (楽曲 2)

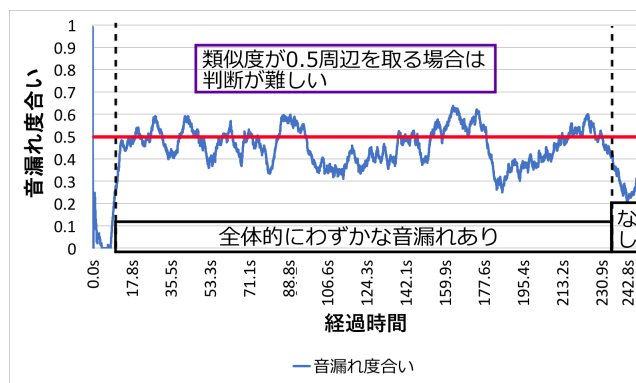


図 20 音漏れモデルを使用した音漏れ度合いの推移 (楽曲 3)

出手法の提案をし、音漏れの検出が可能であると確認した。音漏れ特性のモデル化では sin 波を使い、環境ごとイヤホンごとに音漏れしやすい周波数特性をモデル化した。また、音漏れの検出では音漏れモデルを使い音漏れを仮定し、仮定と実測の音漏れの類似度が音漏れを反映しているとわかった。そして、その類似度を元に定めた音漏れの度合いと主観的評価が概ね一致しているとわかり、音漏れの検出が可能であると確認できた。しかし、より正確な検出には、適切な類似度の上限設定が必要であると考えられる。今後の課題として、主観的な評価ではなく客観的評価で、音漏れの検出精度の調査が必要である。また、音漏れは有無の 2 値でなく、細かな段階分けが必要であると考えられる。加えて、実際のアプリケーションで運用を行う場合は、楽曲を再生してからマイクに観測されるまでの遅延を推定できると望ましい。今後前述の課題を解決し、音漏れの検出のできるアプリケーションを目指す。

参考文献

- [1] 中久治, 土井敦士, 根岸佑也, 梶克彦, “スマートフォンを用いたイヤホン音漏れ測定手法”, 2016, 情報処理学会第 78 回全国大会, 3Q-08.
- [2] 中村一文, 伊藤克巨, “ラベリアマイクを用いた音漏れ検知手法”, 2011, 日本音響学会研究発表会講演論文集 2011,

1-R-18.

- [3] Suzuki, S., Miyabe, S., Kamado, N., et al., “Audio object individual operation and its application to earphone leakage noise reduction”, 3-5 Match (2010), Proceedings of ISCCSP 2010.
- [4] 西村正治, 梶川嘉延, “6 章 アクティブノイズコントロール”, 2011, 電子情報通信学会『知識の森』.
- [5] 西浦敬信, “聴覚マスキングに基づく不快感抑制”, 2014, 日本音響学会誌 70 卷 12 号 (2014), pp.652-657.
- [6] 隅田大地, “環境音による騒音抑制”, 2013, 高知工科大学情報学群 学士学位論文.
- [7] C.M. ビンヨップ 著, 元田浩, 栗田多喜夫 監訳, 樋口知之 他監訳, “パターン認識と機械学習 下 ベイズ理論による統計的予測”, 2007, 丸善出版.
- [8] 石岡恒憲, “x-means 法改良の一提案 —k-means 法の逐次繰り返しとクラスターの再併合—”, 2005, 計算機統計学, Vol.18, No.1, pp.3-13.