

頭部付近のラベリアマイクを用いた ポータブルプレイヤーの音漏れ影響範囲提示

Visualizing Extent of Sound Leakage from Headphones by Using Lavalier Microphones around Head

中村一文

Kazufumi Nakamura

法政大学大学院情報科学研究科情報科学専攻

E-mail: 10t0021@stu.hosei.ac.jp

Abstract

The extent of sound leakage from earphones or headphones must be known when using a portable audio device in order to listen to audio contents in public without disturbing others. We propose a method for detecting sound leakage for visualization. Our method detects sound leakage corresponding to audibility by measuring environmental sounds using a microphone located near a listener's head and a masking threshold. An evaluation experiment showed that the false reject rate was 14% and the false acceptance rate was only 13%, indicating that our method was able to detect most of the sound leakage. And, we estimated extent of sound leakage. The difference of the range of estimate and an actual measurement was less than 15 centimeter in two meter from the result of experimentation.

1 まえがき

本研究では、ヘッドホン聴取時の音漏れの影響範囲を提示する手法を提案する。近年、ポータブルオーディオプレイヤーの普及により、楽曲やラジオ音源等のオーディオコンテンツを屋外で楽しむ人が増加している。しかしながら、同時に音漏れも発生するようになった。音漏れは他人にとって不快に感じるケースが多く、一般に電車やバス内等の公共での音漏れはマナー違反と認識されている。音漏れ問題の難しさはまずプレイヤー使用者が漏出音を自身で聴くことができないため、音漏れしているかどうかを把握できない点にある。また同じレベルの音漏れでも、音漏れは他人との距離によってその影響力が変わる。例えば密着状態で聴こえるぐらいの音漏れは当然それ以上離れれば聴こえない。他人に知覚されないのであれば音漏れとして扱う必要はない。そのため、本研究では音漏れの有無だけでなくその影響範囲も提示することでプレイヤー使用者にとって適切な音量設定を促すことを目標とする。

音漏れの影響範囲を提示するためにはまず音漏れを検出する必要がある。音漏れ発生の第一条件はヘッドホンからの漏出音が環境音にマスクングされないことである。マスクングとは目的音のレベルが雑音をもつ閾値レベル以下の場合に目的音を聞くことができない状態を指す。言い換えると、目的音のレベルが雑音をもつ閾値レベル以上の場合には目的音を聞くことができるといふことである。そのため音漏れの検出にはマスクン

グされないコンテンツの帯域を推定しなければならない。しかしながら漏出帯域はデバイス・オーディオコンテンツ・環境音の組み合わせに依存するため時々刻々と変化してしまう。よって適応的に音漏れを検出するためには、この3つの要素の状態を常に監視することが必要となる。

本研究では頭部付近に設置したラベリアマイクを用いて適応的な音漏れの検出を実現する。また、検出条件にラウドネスを用いたリズム検出 [1] を加えることでより聴感に対応した音漏れ検出を目指す。音漏れの影響範囲も同様にラウドネスのリズムの条件を基に減衰量を算出し、減衰量に届く距離を計算することで聴感に対応した影響範囲を推定する。

2 音漏れの特性と従来対策

2.1 ヘッドホンの音漏れ

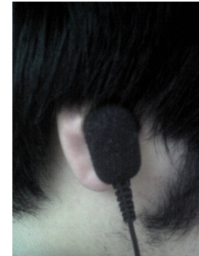


図 1. 漏出音収録例

表 1. Devices

バイノーラルマイク	BME-200(adphox)
プレイヤー	XA-V40-A(victor)
レコーダー	PCM-D50(sony)

ヘッドホン聴取時の音漏れの特性を調べるために、バイノーラルマイクのマイクとヘッドホンを用いて漏出音の音響特性を計測した。図 1 に収録例と表 1 に計測に用いたデバイスをそれぞれ示す。バイノーラルマイクを被験者に装着し、大きな Time-Stretched Pulse(TSP) 信号 [2] をヘッドホンに流し、バイノーラルマイクを用いて漏れた音を収録して計測した 2 名の周波数特性を図 2 に示す。

ヘッドホン聴取時の音漏れの原因として外耳道での共振による特定帯域の強調が挙げられる。この共振は 3-6[kHz] で起こると言われている [3]。今回の計測結果でも 2 名とも異なる共振帯域を持ち、また 3-6[kHz] の帯域内に強調されている成分を確認することができる。さらに再生するプレイヤーの音量を調整し、音圧の異なる TSP 信号を用いて計測した同一被験者内で

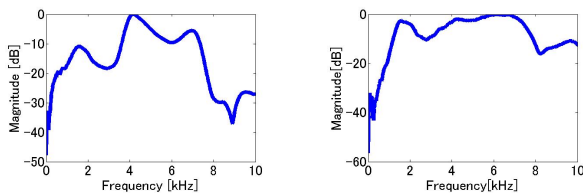


図 2. ヘッドホン聴取時の音漏れ特性

の周波数特性を図 3 に示す．ここで，計測に用いた TSP 信号は音漏れがはっきりとわかる大音量の信号と音漏れがかすかに聞こえる音量の信号の 2 種類である．図 3 より，たとえ音圧が変わったとしても共振の位置を含めて周波数特性が大きく変化することはないことがわかる．そのため，音漏れの特性は音漏れを起こす音圧の範囲内においては個人のなかでその特性が変化することは無い．

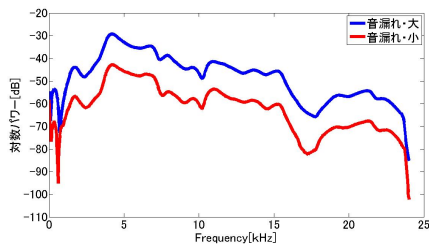


図 3. 音圧の異なる場合の音漏れ特性

次に，実環境においてポータブル音楽プレイヤーで音楽聴取時の音漏れの状況を調べるために，通勤電車の車内でポータブル音楽プレイヤーにバイノーラルマイクを取り付けて音楽を耐えられる最大の音量で再生しながらヘッドホンの周辺の音を収録した．収録された音を聴取して，音漏れを検知できた部分のスペクトルを図 4 に示す．

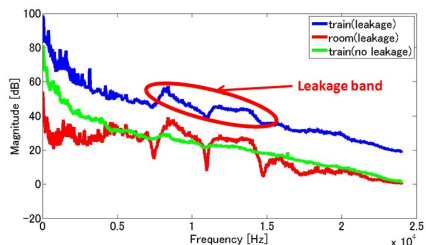


図 4. 音漏れスペクトル

車内で音漏れを発生しながら収録したデータ（青線），雑音が無い環境で音漏れを発生しながら収録したデータ（赤線），漏出音が含まれない車内音を収録したデータ（緑線）のスペクトルを表す．

図 4 の丸で囲んでいる部分が音漏れしている帯域である．この部分は赤線の同じ帯域でのスペクトル形状と似ていることがわかる．このことから音漏れの認知は環境音スペクトルより漏出コンテンツの成分が強く感じられる帯域に起因すると考えられる．つまりは環境音にマスクされない漏出コンテンツの成分が音漏れとして知覚されるということである．

2.2 音漏れ防止機能

これまでこのような音漏れに対するアプリケーションとして AVLS(Automatic Volume Limiter System) やマルチチャンネルオーディオコーデックを利用した漏出音低減手法 [4] が実装・提案されている．

AVLS は，最大ボリュームを予め設定しておき，信号が最大ボリュームを超えないように制御するシステムである．しかし，環境音レベルに適應するものではないので，例えば電車走行中な

どのやかましい場所で設定した最大ボリュームでは，静かな場所では音漏れしやすいという欠点がある．すなわち環境音によるマスキングを考慮していないので，最大ボリュームを設定した環境に依存してしまう．これにより音漏れを防止するためには場所や時間毎に変化する環境音に対して逐一最大ボリュームを設定する必要がある．しかし，そもそもどのような環境でどれくらいのボリュームだと音漏れを起こすのかはこのシステムでは提示されない．したがって，自身の経験や勘で設定せざるを得ないので音漏れを防止するには不十分である．

マルチチャンネルオーディオコーデックを利用した漏出音低減手法は，漏出しやすいチャンネル（音源）の漏出しやすい帯域のみを帯域阻止フィルター（BEF）を用いて低減させ，コンテンツの音質の劣化を最低限に抑えながら音漏れを防止する手法である．BEF を用いる帯域は図 2 で示した外耳道の共振の範囲である 3-6[kHz] である．しかしながら，図 4 で示したように，コンテンツと環境音の組み合わせによっては外耳道の共振の範囲外においても音漏れは発生するため，この手法のように特定の帯域だけを抑制するだけでは音漏れを十分に防止することはできない．

このように，従来の音漏れ防止手法のほとんどは環境音やコンテンツの変化に対して適応的に対応できない．そのため音漏れを高精度で防ぎたいのであれば本研究のようにマイクを用いて常に環境音とコンテンツの変化を監視できることが望ましい．また，本研究は従来研究のようにレベルを抑圧して音漏れを防止するものではなく，あくまで音漏れの有無やその影響範囲を提示するものである．したがって，本研究と従来研究のような抑圧手法を組み合わせることで，音漏れを検出した時点の漏出帯域だけを抑圧して音質の劣化を最小限に抑える等，より効果的な音漏れ対策が可能となると考えられる．

3 頭部周辺のマイクロホンを用いた音漏れ検出手法

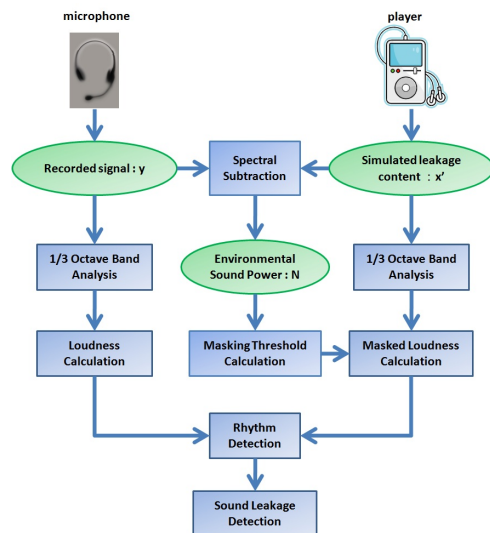


図 5. Detection sound leakage

図 5 に提案する音漏れ検出手法のフローチャートを示す．本研究では音漏れの条件として以下のように設定した：

- 再生コンテンツが環境音にマスキングされない
- 漏出音に対してリズムを感じる

提案手法の構成ではプレイヤーとマイクにより再生コンテンツと漏出音を含む環境音が常に既知となるので，時点ごとに環境音を推定することが可能となる．推定した環境音からマスキング

閾値を求めて再生コンテンツと比較した後、環境音に影響を受けたコンテンツがリズムを有するかラウドネスを算出することで決定する。

3.1 擬似漏出コンテンツ

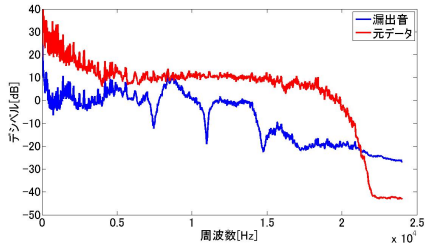


図 6. コンテンツデータと漏出コンテンツデータのスペクトル

マスキング閾値を求めるためにはまず環境音を推定する必要がある。もし、音漏れが発生している場合、頭部付近に設置したマイクによって収録された音 $y(t)$ は、漏出コンテンツ音 $x'(t)$ と環境音 $n(t)$ の和で表現できる：

$$y(t) = x'(t) + n(t) \quad (1)$$

また、マスキング閾値はパワースペクトルを用いて算出されるため環境音パワースペクトル $|N(\omega)|^2$ は以下のように求められる。

$$Y(\omega) = X'(\omega) + N(\omega) \quad (2)$$

$$|N(\omega)|^2 \approx |Y(\omega)|^2 - |X'(\omega)|^2 \quad (3)$$

このように環境音を求めることは可能である。しかしながら漏出コンテンツ $x'(t)$ は図 6 に示すようにデバイスの特性や外耳道共振の影響により元々の再生コンテンツ $x(t)$ とは異なる音響情報をもつ。そのためプレイヤー内のコンテンツ $x(t)$ をそのまま入力として用いることは不可能である。そこで本研究では漏出コンテンツ $x'(t)$ を式 (4) として定義し、図 7 の手順に従って擬似的に漏出データを作成することでこの問題を回避する。

$$x'(t) = x(t) * h(t) \quad (4)$$

ここで $h(t)$ はデバイスや外耳道での共振を含む周波数特性を持つインパルス応答である。この方法はユーザが自分で測定を行う必要があるが、一度測定してしまえばヘッドホンやプレイヤーを変えない限りそのインパルス応答を使用し続けることが可能となるので、ユーザの負担は軽いものとなる。図 8 に作成した擬似漏出コンテンツの周波数特性を示す。図 6 の本来の漏出コンテンツの特性と比べるとよく近似できていることがわかる。

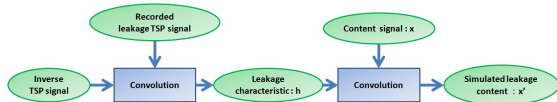


図 7. Simulated leakage contents

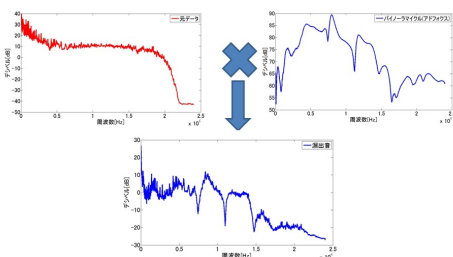


図 8. 擬似漏出データ作成

3.2 マスキング閾値

マスキング閾値は前述で求めた環境音パワースペクトル $|N(\omega)|^2$ から計算できる。マスキング閾値を計算するためのステップを以下に示す [5]：

- クリティカルバンドエネルギー (図 9-左上): クリティカルバンド毎のパワースペクトルの総和
- 拡散クリティカルバンドスペクトル (図 9-右上): クリティカルバンドエネルギーと拡散関数の畳み込み
- 拡散閾値 (図 9-左下): 拡散クリティカルバンドスペクトルから閾値オフセットを減算
- 最終閾値 (図 9-右下): 絶対聴覚閾値と拡散閾値を比較

もしコンテンツのレベルがある帯域で環境音の最終閾値を上回るなら、その帯域で漏出は発生することになる。

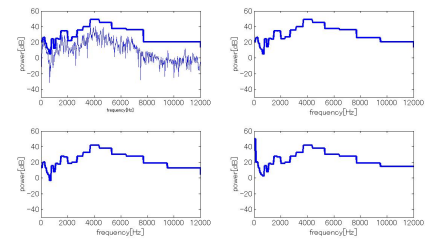


図 9. masking threshold

3.3 リズム検出

コンテンツのパワーがマスキング閾値を上回る場合に漏出は発生する。しかしながら、その音が音漏れとして他人に知覚されるかはわからない。例えば、漏出音がホワイトノイズのような定常的な音の場合、それは環境音と区別できない可能性がある。このような場合では他人がその音を気にならないので、音漏れとして検出する必要がない。本研究では音漏れとして気になる音をリズムを有する音として捉える。

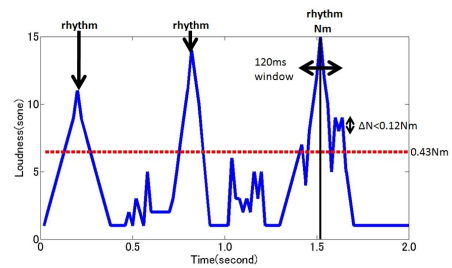


図 10. model of rhythm

図 10 にラウドネスによるリズムのモデル [1] を示す。ラウドネスとは物理的な音の強さではなく、人間が感じる音の大きさを表す心理量である。人間は周波数毎に物理的な音の強さと感じる音の大きさの関係が異なるため、ラウドネス曲線を考慮して補正する必要がある。ラウドネスと音の強さの関係は以下のようにベキ関数で表現できる [6]：

$$N = kI^\alpha \quad (5)$$

ここで N はラウドネス、 I は音の強さ、 α と k は定数であり帯域によって異なる。ラウドネスは人間が感じる音の大きさをよく表しており、より主観的な尺度と言える。このラウドネスを用いることで人が感じるリズムは以下の条件を満たすことで検出できる：

$$\begin{cases} N > 0.43Nm \\ \Delta N > 0.12Nm \\ N \text{ が前後 } 120ms \text{ 内での最大値} \end{cases} \quad (6)$$

ここで Nm はある区間 (例えば音楽の 1 フレーズ) でのラウドネスの最大ピーク, ΔN は前フレームとの差分である. これらの条件により図 10 の例では 3 つのリズムを検出することができる.

このモデルはある 1 種類の音におけるリズムのモデルである. しかしながら, 今回対象とするのは環境音下で漏出音がリズムを有するかを検出するため, このモデルをそのまま用いることは難しい. とところで漏出音は環境音の影響を受けるため基本的に本来のラウドネスより小さくなる. この状態のラウドネスをマスクラウドネスと呼ぶ. 本研究では環境音下での漏出音のリズムを検出するために, 式 (6) において漏出音のマスクラウドネスを N ・環境音のラウドネスのピークの平均を Nm と置き換えたモデルを提案する. またリズムは 1/3 オクターブバンドごとに検出する. これは人間が臨界帯域毎に音の大きさを感知できるため, リズムを感じる音の大きさもまた臨界帯域毎に存在するためである.

3.4 雑音下におけるリズム検出

雑音が存在する場合の目的音のラウドネス (マスクラウドネス) は以下の式で求めることができる [7]:

$$N = k(I_t^\alpha - I_{thr}^\alpha) \quad (7)$$

ここで, I_t は目的音の音の強さ, I_{thr} が雑音によって上昇したマスクング閾値である. 目的音が複合音の場合, 帯域毎にマスクラウドネスを求める. しかしながらセクション 3.2 で求められたマスクング閾値はクリティカルバンド毎に求められたものに対し, マスクラウドネスは 1/3 オクターブバンド毎に求めなければならない. そのため先のマスクング閾値を 1/3 オクターブバンド用に補正する必要がある. 補正式は以下になる [8]:

$$L_b = L_S + 10 \log \frac{W_{1/3}}{W_{CB}} \quad (8)$$

ここで L_b は補正された閾値レベル, L_S は閾値レベル, $W_{1/3}$ は 1/3 オクターブバンド幅, W_{CB} はクリティカルバンド幅である. 補正された閾値を用いてマスクラウドネスを式 (7) で求め, リズムの条件式 (6) に当てはめて検出を行う. しかしながら, 雑音下においてリズムの条件のひとつである $\Delta N > 0.12Nm$ は適用できないため本研究ではこの条件を削除した. そのため, 最終的なリズムの条件は以下ようになる.

$$\begin{cases} N > 0.43Nm \\ N \text{ が前後 } 120ms \text{ 内での最大値} \end{cases} \quad (9)$$

ここで N が漏出音のマスクラウドネス, Nm が漏出音を含めた環境音のラウドネスのピークの平均である.

3.5 評価実験

評価実験はあらかじめ正解を付けた評価データと手法によって判定された検出結果を照らし合わせ, FAR・FRR をそれぞれ求めて評価した. ここで, FAR は, 「実際には音漏れが起きていないが, 音漏れであると誤判定された割合」, FRR は「実際には音漏れが起きているが, 音漏れではないと誤判定された割合」である.

評価データは実環境下でバイノーラルマイク・ヘッドホンを用いて音楽プレイヤーからコンテンツを再生させながら収録したデータである. 収録は自室, 路上, 電車内, コンビニ等普段の生活でポータブルプレイヤーを活用する場で行った. 収録時間は 90 分, ここからランダムに抜き出した 200 個の 5 秒間のデータに対し, 耳で聴いて音漏れの正解を付けた (音漏れ有: 103, 音漏れ無: 97).

実験では以下の 3 種類の手法をそれぞれ評価した:

1. リズム検出を行わず, マスキング閾値のみで音漏れを判定
2. 従来の Nm にラウドネスの最大ピークを用いたリズム検出による音漏れを判定
3. 提案した Nm にラウドネスのピークの平均を用いたリズム検出による音漏れ判定

ここで 2,3 の手法で用いる Nm は過去 1.5 秒間の区間で求めている. 1 では 5 秒間のうちに漏出と判定されたフレーム数を閾値として閾値ごとの FAR, FRR を算出し, また 2,3 では 5 秒間のうちにリズムが検出された数を閾値として閾値ごとの FAR, FRR を算出し各手法の ROC 曲線を求めた. 表 2 に実験条件を示す.

表 2. 実験条件

サンプリング周波数	48000[Hz]
量子化ビット数	16[bit]
FFT 幅	2048 点
シフト幅	1024 点
1/3oct. バンド幅	62.5-12,500[Hz](24bands)

3.6 結果と考察

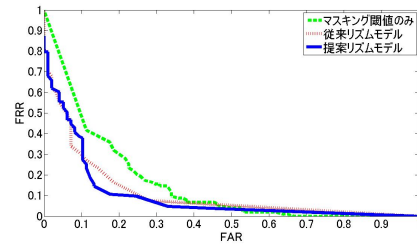


図 11. 検出結果 (ROC 曲線)

表 3. 検出結果 (best performance)

	手法 1	手法 2	手法 3
FAR	39%	18%	13%
FRR	4%	16%	14%

図 11 に ROC 曲線による実験結果を示す. また表 3 に各手法で総合的な性能が最も良かった場合の FAR, FRR をそれぞれ示す. 表 3 より手法 1 (リズム検出をしない, マスキング閾値のみ) では FAR が 39% となり, 過剰検出が目立った. これは音が漏れていることだけが音漏れとして人に認識される要因ではないことを表している. そしてリズムのモデルを用いた手法 2 では FAR が 18% となり, このことからリズムを感じることが音漏れの認識の要因に関連していると考えられる. しかしながら, 手法 2 では FRR が 16% となり, 誤検出が目立った. これはリズムのモデルが雑音 (環境音) の存在を加味していないモデルであるからと考えられる. コンテンツは比較的一定のピークの大きさをもつにも関わらず, 雑音が付加されるとピークの大きさが雑音の分だけ変動してしまう. そのためリズム判定のための閾値がそのままでは使用できない可能性がある. 一方, 提案手法 (手法 3) はピークの平均をとることにより雑音の

パワーの平均（時間変動しない値）をコンテンツに重畳させることができる。そのため提案手法ではリズム判定のための閾値をそのまま用いることが可能となる。結果として、提案手法は FAR が 13 %, FRR が 14 % となり、安定した検出を実現することができた。

今回、提案手法では 5 秒間のうちに検出されたリズムの数を閾値として音漏れを判定していた。しかしながら、5 秒間のうち局所的にしか漏出していない音の場合、5 秒間のうちに存在するリズムの個数は当然少ない。よって単純に 5 秒間のリズムの個数を閾値にすると、閾値に設定した個数が少ないと FAR より、個数が多いと FRR よりになってしまう。この問題を解決し、更に音漏れの検出精度を向上させるためには、推定した漏出音のレベルが高いと区間内のリズムの個数が多くなり、またレベルが低いとリズムの個数は少なくなるはずなので、それらを基に閾値を適時変化させていくことが必要だと考えられる。また、環境音のレベルが高い場合には観測されるリズムのピーク値の分散も大きくなるので、ピークの分散が大きい場合には閾値とするリズムの個数を少なく、分散が小さい場合には閾値とするリズムの個数を多くするなどの工夫が必要である。

4 音漏れ影響範囲推定

4.1 距離減衰

音漏れの影響範囲は、言い換えれば音漏れを感じなくなる閾値を下回るまで減衰した場合の距離である。そのため本研究では音の距離減衰を用いて音漏れの影響範囲を推定する。距離減衰は以下の式で表すことができる：

$$L_p = L_w + 10 \log \left(\frac{Q}{4\pi r^2} + \frac{4}{A} \right) \quad (10)$$

ここで L_p は音源からの距離 r [m] における音圧 [dB]、 L_w は音源の強さ [dB]、 Q は指向係数、 A は空間の広さと素材による吸音作用 [m^2] である。指向係数 Q は音源の位置が自由空間（空中に浮いている）で 1、半自由空間（地面に置いてある）で 2、1/4 自由空間（壁際）で 4、1/8 自由空間（部屋の隅）で 8 となる。また吸音作用 A は以下の式で表現できる：

$$A = \bar{\alpha} \times S \quad (11)$$

ここで $\bar{\alpha}$ は平均吸音率、 S は部屋の表面積となる。

音漏れはリズムを感じることで知覚されるとすると、音漏れの影響範囲は、漏出音がリズムのモデルの条件に当てはまらない、つまり $N \leq 0.43Nm$ になるまで減衰する距離を求めればよい。しかしながら、 Nm は漏出音と環境音を含めた音から求めているため、音源から距離が離れると漏出音の影響が弱くなり Nm の取り得る値が変化してしまう。そのため、この範囲推定では Nm を漏出音を含まない場合の環境音のピークの平均値とする。式 (5) を用いて N と Nm を変形すると $N \leq 0.43Nm$ となるための減衰量 [dB] は以下となる：

$$\text{減衰量 [dB]} = 10 \log_{10} \left(\frac{N}{k} \right)^{\frac{1}{\alpha}} - 10 \log_{10} \left(\frac{0.43Nm}{k} \right)^{\frac{1}{\alpha}} \quad (12)$$

したがって影響範囲（減衰距離） r_{ex} は：

$$r_{ex} = \sqrt{\frac{Q}{4\pi \left(10^{\frac{-減衰量}{10}} - \frac{4}{A} \right)}} \quad (13)$$

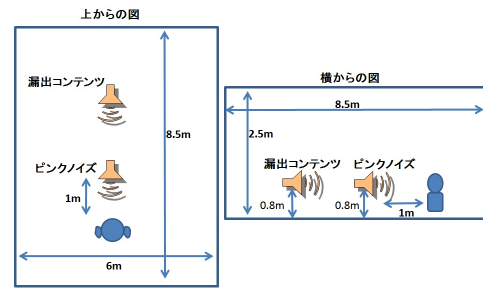


図 12. 範囲推定実験の様子

表 4. 実験条件

指向係数:Q	2
部屋の表面積:S	175.5[m ²]
平均吸音率: $\bar{\alpha}$	0.33
スピーカー	Mediamate(BOSE)
プレイヤー	XA-V40-A(victor)
被験者数	6人

4.2 実験

実験は推定した距離に対して実際に音漏れを感じる距離を主観で評価してもらうことで行った。図 12 のように部屋の中に 1 組のスピーカーを置き、片方のスピーカーからはピンクノイズ、もう片方のスピーカーからは漏出特性を付加させたコンテンツ音を再生する。表 4 に実験条件を示す。部屋の表面積は実験前に予め測定した。また平均吸音率も事前に TSP 信号を用いて測定した。平均吸音率の算出手順はまず図 12 の被験者が座る位置にマイクを置き、スピーカーから再生した TSP 信号を録音した。その後、録音した TSP 信号に対し逆 TSP 信号を畳み込み得られたインパルス信号から残響時間 (s) を見積もった。ここで残響時間はインパルス信号のパワーが 60[dB] 減衰するまでの時間とした。最後に残響時間から平均吸音率を求めするために Sabine の残響式を用いた：

$$T = \frac{0.161V}{\bar{\alpha}S} \quad (14)$$

$$\bar{\alpha} = \frac{0.161V}{TS} \quad (15)$$

ここで $\bar{\alpha}$ は平均吸音率、 T が残響時間 (s)、 S が部屋の表面積 (m^2) 及び V が部屋の体積 (m^3) である。実験条件より Q と A を定めた場合の距離と減衰の関係を図 13 に示す。図 13 より 2m 以降では減衰率がだんだんと収束していくのがわかる。このことから今回の実験では 3m までの推定距離を評価する。実験では空間内の環境音レベルは一定と仮定して、ピンクノイズを再生するスピーカーは固定し、コンテンツ音を再生しているスピーカーの位置だけを変えて影響範囲を測定した。実験は被験者に対して最初からコンテンツ音を感じない距離までコンテンツのスピーカーを離しておき、その後コンテンツのスピーカーをだんだんと被験者に近づけていき、被験者が初めてコンテンツ音を感じる距離で手を挙げてもらうことで音漏れの範囲を測定した。各被験者に対して測定を二回行い、測定に使用するコンテンツも二つ用意した。そして全ての測定結果の平均を主観評価による音漏れの影響範囲とした。なお実験においてスピーカーは 25cm ずつ移動させた。

4.3 結果と考察

図 14 より、推定距離 2[m] までにおいて実際に音漏れと感じる平均距離との誤差が 15[cm] 以内となり精度のよい結果と

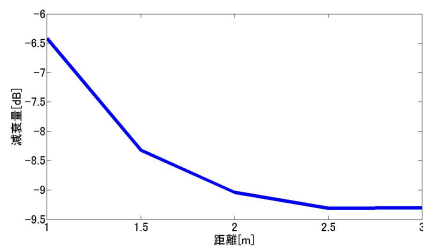


図 13. 距離と減衰量の関係

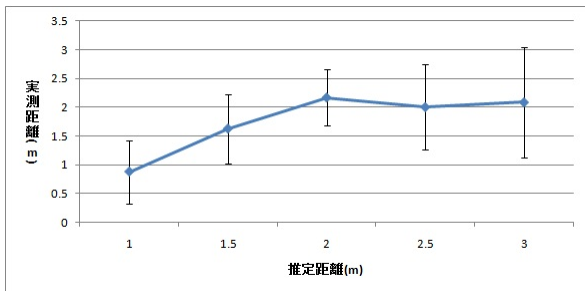


図 14. 範囲推定実験結果

なった。しかしながら、推定距離 2.5[m] と 3[m] では実測の距離との誤差が大きく、推定距離 2.5, 3[m] とともに実測距離の平均が 2[m] 付近となった。これは図 13 より音の減衰が 2[m] 以降で収束しだしていることに起因しているものだと考えられる。また人によって音漏れを感じる距離にばらつきがあり、これは今回の実験における音漏れの基準が被験者間で異なっていたことが原因と考えられる。今回の実験では被験者の音漏れだと判断する基準は主に以下の 3 つだった：

1. 「一瞬でもノイズとは異なる音が聴こえる」
2. 「何の音かまではわからないがコンテンツのような音が聴こえる」
3. 「コンテンツの存在が認識できるまで聴こえる」

1 のような判断ができた理由としては今回の実験で用いたピンクノイズが時間的に変化しない音であったので、そこに一瞬でも異音があれば気付やすいことである。被験者からも「普段は音漏れとして気付かないが、ノイズがなにかがわかっているので一瞬でも鳴っている音が違うとわかってしまう」とあり、このような場合は推定距離よりも遠い距離で音漏れとして判断されることが多かった。2 の場合では推定距離と近い距離で音漏れと判断されることが多く、リズムを感じ取ることで音漏れを判断していると考えられる。3 の場合はコンテンツがより明瞭に聴こえる距離まで近づけるため、推定距離と比べて近い距離で音漏れと判断されることが多かった。以上のことより、今回の手法は 2 のような判断で音漏れを決定する場合には有効であることがわかる。これは言い換えれば音漏れの基準によって有効な手法はそれぞれ異なるということである。例えば、1 の場合に対応するにはリズムを用いずマスキング閾値だけで音漏れの判断をし、3 の場合ではより再生されているコンテンツに近い音漏れをを探索するため音色類似度を用いて判断することが考えられる。

今回の実験ではあらかじめ空間の大きさなどを調べてパラメータとして減衰式に当てはめたが、実際にアプリケーションとして使用する場面では、マイクロホンから得られる音響情報のみで空間情報を入力するのは困難である。しかしながら、空間の変化は音の変化と比べて激しいものではないと考えられる。例えば電車の中では、音の大きさは電車の走行速度や、ア

ナウンスや乗客の話し声などで時々刻々と変化するのに対して、電車という箱の大きさは時間によらず常に一定である。そのため、ある程度場所を分類し、分類ごとの空間や吸収作用のパラメータをプリセットとして持っておけば、プレイヤーの操作回数は最低限で済むと考えられる。

5 あとがき

ポータブルプレイヤー使用時の音漏れを適応的に検出しその影響範囲を提示するために、ラベリアマイクを用いた手法を提案した。音漏れの知覚の要因をリズムを感じることに定義し、ラウドネスによるリズムのモデルを用いて検出と範囲推定を行った。

検出部では、リズムのモデルを雑音下でも適応できるように改めた結果、検出実験により FAR が 13%, FRR が 14% となり安定した検出を実現することができた。今後の課題としては、今回検出区間内のリズムの個数を閾値として音漏れを判定していたため、漏出音のレベルや背景音のレベルによって音漏れと判定するためのリズムの閾値を適応的に変化させることが挙げられる。

範囲推定部では、リズムの条件から外れるまでに必要な減衰量から距離を推定した。聴取実験では推定距離 2m までにおいて実測した平均距離との誤差が少なかった。これは減衰量が収束しだす前の距離までは精度よく推定できることを表しており、部屋の大きさや吸音率によって手法が有効な範囲は変化すると考えられる。また被験者によって音漏れを感じる距離が大きく異なり、これは被験者間で音漏れの基準が異なることが原因と考えられる。そのため今後は音漏れの基準ごとに手法を設定し実験する必要がある。

検出部と範囲推定部の共通の課題として、今回は音漏れの正解付けにヘッドホン聴取や範囲推定のために雑音空間を作るなど、擬似的な実験のみ行ったため、今後は電車の中などの実際の環境下での性能を試す必要がある。

参考文献

- [1] “PSYCHOACOUSTICS Third Edition” Hugo Fastl Eberhard Zwicker, pp271-276, Springer, (2006)
- [2] Y. Suzuki, F. Asano, H.-Y. Kim, and Toshio Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses” J. Acoust. Soc. Am. Vol.97(2), pp. 1119-1123 (1995)
- [3] “The Science of SOUND Third Edition” Rossing, Moore Wheeler, pp.81, Pearson Education, (2002) .
- [4] S. Suzuki, et.al “Audio object individual operation and its application to earphone leakage noise reduction” in Proc. ISCCSP 2010, 3-5 March (2010).
- [5] James. D. Johnston, “Transform Coding of Audio Signals Using Perceptual Noise Criteria” IEEE J. on Select. Areas Commun., vol. 6 pp. 314-323. February (1988)
- [6] S. S. Stevens, “Perceived Level of Noise by Mark VII and Decibels (E)” J. Acoust. Soc. Am. 51, pp. 575-601. (1971)
- [7] J. P. A. Lochner and J. F. Burger, “Form of the loudness function in the presence of masking noise” J. Acoust. Soc. Am. 33, pp. 1705-1707. (1961)
- [8] Y. Suzuki, et.al “The optimum level of music listened to in the presence of noise ” J. Acoust. Soc. Jpn. 3, pp. 55-65. (1982)