

推薦論文

ウェアラブルコンピューティングにおける 周波数操作による聴力自在化技術の提案

渡邊 拓貴^{1,a)} 寺田 努^{2,b)}

受付日 2020年4月17日, 採録日 2020年11月5日

概要: 人の聴力は自分で制御することが難しく, 必要のない情報まで取得したり, 重要な情報を逸失したりすることがある. 本研究では, マイクとスピーカを搭載したイヤホン型のウェアラブルデバイス (ウェアラブルデバイス) に着目した. ヒアラブルコンピューティング環境では, ユーザはつねにマイクとスピーカを耳に装着しており, ユーザの聴力を自在に操作できる環境だといえる. そこで本研究では, 外界音を変換し, ユーザが自在に自身の聴力を操作するための聴力自在化技術のフレームワークを提案する. 具体的には, マイクで取得した外界音の周波数を操作し, 変換後の音をスピーカでユーザに提示することで, 従来の聴力では聞こえなかった音の取得や, 不必要な音の削除が可能となる. 本研究では, 周波数の操作方法として5種類の方法を提案し, プロトタイプシステムを実装した. さらに, 聴力の自在化によって実現できる7種類の想定アプリケーションを提案した.

キーワード: ウェアラブルコンピューティング, 人間拡張, 聴力, 超音波

Manipulatable Auditory Perception using Frequency Manipulation in Wearable Computing

HIROKI WATANABE^{1,a)} TSUTOMU TERADA^{2,b)}

Received: April 17, 2020, Accepted: November 5, 2020

Abstract: We proposed a framework to manipulate auditory perception. Since auditory perception is passive sense, we often acquire unimportant information and do not notice important information. In this study, we focused on an earphone-type wearable computer (hearable device) that not only has speakers but also microphones. In a hearable computing environment, we always attach microphones and speakers to the ears. Therefore, we can manipulate our auditory perception using a hearable device. We manipulated the frequency of the input sound from the microphones and transmitted the converted sound through the speakers. Thus, we could acquire the sound that is not heard with our normal auditory perception and eliminate the unwanted sound according to the user's requirements. We devised five types of frequency-manipulating techniques and implemented a prototype device. Moreover, we proposed seven assumed applications that can be realized by the proposed framework.

Keywords: Wearable computing, augmented human, auditory perception, ultrasound

1. はじめに

近年のコンピュータの小型化にともない, ウェアラブル

コンピュータが一般的になりつつある. すでに多数の頭部装着型ディスプレイ (HMD: Head Mounted Display) やスマートウォッチが市販されており, ウェアラブルコンピュータを利用できる環境にある. ウェアラブルコンピューティング環境では, デバイスをつねに装着している

¹ 北海道大学
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan

² 神戸大学
Kobe University, Kobe, Hyogo 657-8501, Japan

^{a)} hiroki.watanabe@ist.hokudai.ac.jp

^{b)} tsutomu@eedept.kobe-u.ac.jp

本論文の内容は2020年3月のインタラクシオン2020にて報告され, 同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である.

ため、デバイスによってユーザの能力を拡張/制限することが可能になる。たとえば、視覚の拡張/制限が研究されており、HMDを用いた視覚障害のサポート [1]、視野の拡張 [2]、視覚の制限 [3] などが提案されている。特に近年注目を集めるウェアラブルデバイスとして、ヒアラブルデバイスがあげられる。ヒアラブルデバイスはイヤホン型のウェアラブルコンピュータであり、従来のイヤホンの用途である音楽鑑賞だけでなく、スマートフォンと連携することによる音声アシスタントの利用や、通話なども可能である。そのため、多くのヒアラブルデバイスにはスピーカだけでなく、音声認識や通話に用いられるためのマイクが内蔵されている。したがってヒアラブル環境では、ユーザはつねに耳にマイクとスピーカを装着している環境だといえる。本研究では、人間の五感の中でも特に重要な聴力に着目する。人間の聴力は全方向からの音を取得でき、視覚外からの情報の取得に特に役立っている。人間が聞くことができる音はおよそ 20 Hz から 20 kHz といわれており、多くの環境音は 20 kHz 以内に存在するが、電気自動車 (EV: Electric Vehicle) やハイブリッド車 (HV: Hybrid Vehicle) の走行音 [4]、不調時の機械音 [5]、気体の漏洩時 [6] など、20 kHz 以上の領域 (超音波領域) にも重要な音が存在することが知られている。これらの超音波領域に含まれる重要な音を知覚することができれば、思わぬ事故を避けられる可能性がある。上記したような、耳に常時デバイスを装着しているヒアラブル環境では、ユーザの要求に応じて不要な音の削除や、超音波の可聴化など、人の聴力を自由自在に操作できる環境だといえる。

これらの目標を念頭に置き、本研究では外界音を変換し、ユーザが自在に自身の聴力を操作するための聴力自在化技術のフレームワークを提案する。特に、本論文では周波数操作による聴力の自在化に着目した。周波数操作の手法として、compression, slide, overlap, partial elimination, partial extraction の 5 種類を提案した。さらに、聴力自在化によって実現可能なアプリケーションを 7 種類提案した。実装したプロトタイプデバイスを用いた調査の結果、周波数操作後においても音源方向の検出が可能ことが確認できた。また、周波数操作にともなう音の遅延と音質のトレードオフについても調査を行った。なお、本論文は文献 [7] で発表した内容に議論を追加し、全体的に補完を行ったものである。

以降、2 章で関連研究について述べ、3 章でフレームワークを実現するための提案手法について説明する。4 章で実装について述べ、5 章で提案フレームワークで実現できる想定アプリケーションを紹介する。6 章で提案フレームワークの評価を行い、7 章で考察を行う。最後に 8 章で本論文をまとめる。

2. 関連研究

2.1 外界音の変換

外界音を変化させることでユーザの聴覚を拡張する研究や製品は複数あり、これらの関連研究が本研究に最も近い。コウモリ探知機は超音波を可聴化するデバイスであり、コウモリの発信した超音波を人が聞こえる領域に変換することで、コウモリの調査に用いられる [8]。補聴器の Sound Shaper は、入力周波数と出力周波数の比例関係を維持しながら、高周波音を聞こえる領域に移動できる [9]。Speech Rescue は、聞き取れない高周波音をコピーし、可聴域に変換する周波数合成手法である [10]。Here One はワイヤレスイヤホンであり、音楽のストリーミング、音声の増幅などができる [11]。また、電話をかけることや、ノイズなどの特定の音をフィルタリングすることもできる。Bose は、外界音のイコライジングや、取得する音の方向を変化できるデバイスである Hearphones を発売している [12]。Tajadura-Jiménez らは、歩行音を変化させることで自分の知覚する体重を変化させ、それに関連する歩行パターンになることを示した [13]。彼らは、さらに足音の変化で知覚する男らしさや女らしさが変化するということまで調査している [14]。エコーロケーションは、コウモリやイルカなどの動物が使用するソナーである。超音波を発信し、周辺の物体からのエコーを利用して物体の位置や種類を識別している。この超音波を可聴化することで、人によるエコーロケーション (ヒューマンエコーロケーション) の可能性を調査する研究が行われている [15], [16], [17]。Weger らは、音のコントラストをリアルタイムで強調することにより、音を聞きやすくする手法を提案している [18]。Albrecht らは、バイノーラルイヤホン装着時に生じる外耳道内の共鳴を補正し、外界音を自然に取得できる手法を提案した [19]。Storek らは外界音に対して様々なサウンドエフェクトを付加する手法を提案した [20]。この手法は我々の提案に近いものであるが、本研究では特に周波数の操作に着目し、既存のサウンドエフェクトでは用いられない、実環境での利用を考慮した周波数操作手法を提案した。

本研究は、これらの先行研究に基づいて提案されている。本研究では任意の周波数を操作できる聴力自在化技術のフレームワークを提案しており、これらの関連研究で行われていることは、本研究に包含されているという位置付けである。今後行われる外界音を変換してユーザに提示する研究に、本研究のフレームワークが利用できると考える。

2.2 ヒアラブルデバイスを用いたインタラクション

ヒアラブルデバイスを用いたインタラクションについても幅広く研究が行われている。Bedri らは、赤外線近接センサを用いて下顎の動きにともなう外耳道の変形を取得し、インタフェースとして利用する手法を提案した [21]。

また、彼らはイヤホン型デバイスを用いて食事活動の認識も行っている [22]. CanalSense は、顎、顔、頭の動きを、イヤホンに埋め込まれた気圧センサを利用して認識するシステムである [23]. Manabe らは、導電性ゴムでできたイヤチップにより眼電位を計測し利用するアイジェスチャインタフェースを提案した [24]. Laput らは、イヤホンから発信した超音波スイープ信号の応答を利用し、ユーザーの状況を認識する手法を提案した [25]. Arakawa らは、外耳道の音響特性を利用して個人認証を行う手法を示した [26]. NEC は、ヒアラブルデバイスに内蔵された地磁気センサを用いた屋内測位技術を開発している [27]. Amesaka らは、外耳道の伝達関数を利用することでユーザーの顔の動作を認識している [28].

ヒアラブルデバイスを用いたインタラクション技術に関する研究は多く行われているが、外界音の周波数を自在に操作する研究は我々の知る限り行われていない。また、これらの研究ではデバイスがコンテキストを認識する。提案フレームワークでは、デバイスによって聴覚を操作するが、その音によってコンテキストを判断するのはユーザー自身である。

2.3 超音波を利用したシステム

従来の聴力では、多くの人が超音波を知覚することはできないが、特定の場面では超音波領域にも情報があることが知られている。Takagi らは、EV/HV から生成される高周波音によって、走行音が静かな EV/HV を検知できる手法を提案した [4]. 3DSignals は、機器の近くに取り付けられた超音波センサモジュールによって機器の異常時に発生する超音波を検出し、異常を検知できるシステムである [5]. 気体の漏洩時にも超音波が発生することが知られており、異常検出の手法として用いられている [6].

これらの研究では、システムが超音波を取得、解析し、コンテキストを認識している。本研究の目標は、周波数を編集することで人の聴覚を操作することである。操作された聴覚により得られた情報を用いて、コンテキストを認識するのはユーザー自身である。

3. 提案手法

3.1 システム構成

本研究では、周波数変換手法としてフェーズボコーダを用いる [29]. 一般的に、周波数スペクトルを計算した後に、周波数領域でパワースペクトルを編集する。この手法により、ユーザーは周波数スペクトル上で直感的に周波数を操作できる。周波数変換の手法として、時間伸長やヘテロダインなどの手法 [8] も考えられるが、これらの手法では詳細な周波数操作が難しい。提案フレームワークではあらゆる周波数を自在に操作できる環境を考えているため、フェーズボコーダが最適であると考えた。しかし、フェーズボコー

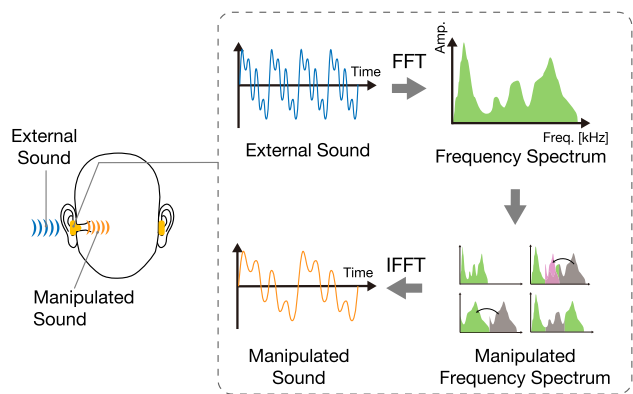


図 1 システム構成

Fig. 1 System configuration.

ダは音声劣化を完全になくすのは難しい点や、計算時間がかかる点など、いくつかの欠点もある [29]. これらの問題点の改善は今後の課題とする。

図 1 にシステム構成を示す。本研究では、ユーザーはマイクが内蔵されたヒアラブルデバイスを装着している環境を想定する。まず、ヒアラブルデバイス内蔵のマイクで外界音を取得する。取得された音は高速フーリエ変換 (FFT: Fast Fourier Transform) によって周波数領域に変換され、3.2 節で述べる周波数操作手法に基づいて周波数領域で操作される。次に、操作された周波数スペクトルは逆高速フーリエ変換 (IFFT: Inverse Fast Fourier Transform) によって時間領域に変換され、ユーザーはスピーカから発信される操作された音を取得する。ユーザーは、周波数操作手法と、その適用範囲をデバイス上で操作することで、自身の聴力を自在に操作できる。本研究では、マイクからの録音とスピーカからの送信のサンプリング周波数は 96 kHz とした。FFT のサンプル数は 1,024 とし、512 サンプルずつオーバーラップさせた。FFT を計算する際には、周波数分解能と時間分解能はトレードオフの関係にある。つまり、FFT サイズを大きくすれば周波数領域を詳細に操作できるが、遅延時間が大きくなる。一方、FFT サイズを小さくすれば遅延が少なくなるが、周波数の分解能が下がる。本研究では、周波数分解能と遅延時間を考慮して、FFT のサンプル数を 1,024 とした。詳細な議論は 7.1 節で行う。

3.2 周波数操作手法

図 2 に示すように、5 種類の周波数操作手法を提案する。周波数領域で操作することにより、ユーザーは聴力を自在に操作できる。提案フレームワークの重要な特徴の 1 つは、ある周波数を別の周波数に自在に移動させることができる点にある。この特徴を考慮して、5 つの周波数操作手法を提案する。

3.2.1 Compression

compression は、図 2 (a) に示すように、全周波数を可聴域に圧縮する手法である。位相情報も同様に圧縮される。

操作後の最大周波数を操作することで、圧縮率を変化させることができる。compression では全周波数の位置関係が保持されるため、複数の音の関係が重要なときに利用できる。一方、全周波数を圧縮するため、可聴域の音声も変化する。

3.2.2 Overlap

overlap では、非可聴域の音を可聴域に重ねる操作を行う (図 2(b))。本手法は Speech Rescure [10] で用いられているものに近いが、超音波を含む任意の範囲の音を、任意の範囲に重畳させることができる。本手法では、従来の可聴音と、通常の聴力では聞くことのできない超音波を同時に取得できる。したがって、本手法は混雑した場所や道路のような、従来の可聴音がなくなると危険な場面での利用を想定している。

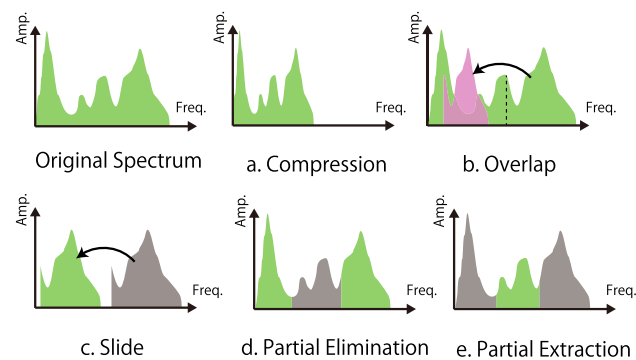


図 2 周波数操作手法：(a) compression, (b) overlap, (c) slide, (d) partial elimination, (e) partial extraction

Fig. 2 Frequency-manipulating techniques: (a) compression, (b) overlap, (c) slide, (d) partial elimination, and (e) partial extraction.

3.2.3 Slide

slide では、周波数領域の一部を別の周波数領域に移動させることができる (図 2(c))。位相情報も同様に移動させる。slide では、ある特定の領域にだけ着目できる。たとえば、超音波領域を可聴域にスライドすることで、超音波だけに着目することや、特定の周波数を拡大して小さな違いを聞き取る場合にも利用できる。

3.2.4 Partial Elimination

partial elimination は、図 2(d) に示されるように、可聴域の一部分のみを削除する操作手法である。ユーザは、自身に必要な周波数領域を選択することで、必要のない周波数領域の情報だけを削除できる。これはバンドストップフィルタと同様のものといえる。

3.2.5 Partial Extraction

partial extraction は、図 2(e) に示されるように、可聴域の一部分のみを抽出する操作手法である。ユーザは、自身に必要な周波数帯域を選択することで、必要な周波数領域の情報だけを得られる。これはバンドパスフィルタと同様のものといえる。

4. 実装

4.1 ハードウェア

実装したプロトタイプを図 3 に示す。プロトタイプは、タブレット端末 (Huawei MediaPad M5)、オーディオインタフェース (Zoom U-24)、イヤホン (Shure SE215)、マイク (Countryman B6) から構成されている。我々の調査した範囲では、オーディオインタフェースを介して 96 kHz でサンプリングできるデバイスが Mediapad M5 のみであったため、本論文ではこのデバイスを利用した。可能な限り

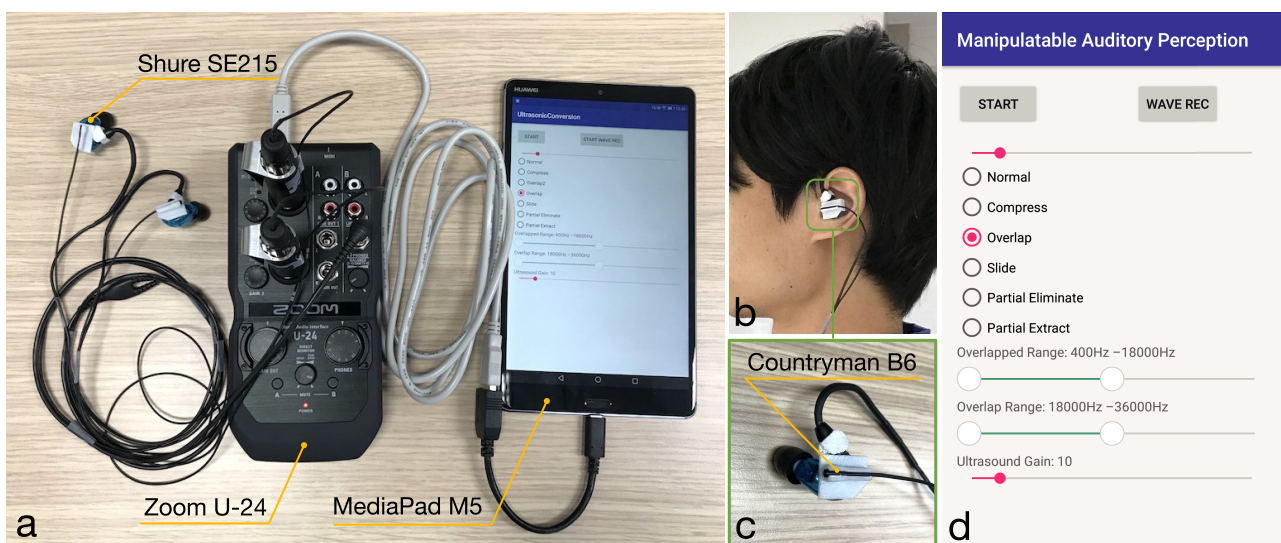


図 3 (a) 実装したプロトタイプデバイス, (b) デバイス装着時, (c) イヤホンとマイクの詳細な外観, (d) 実装した Android アプリケーション

Fig. 3 (a) Implementation of the prototype, (b) wearing the device, (c) detailed appearance of the earphone and microphone, and (d) screenshot of the application.



図 4 Xperia XZ1 を用いて実装したプロトタイプ
Fig. 4 Implemented prototype using Xperia XZ1.

外界音を遮断し、変換後の音のみをユーザが取得するため、イヤホンには高遮音性に優れた Shure SE215 を用いた。マイクは、小型でイヤホンに装着できる Countryman B6 を用いた。このマイクの周波数特性は 30 Hz から 20 kHz となっているが、予備実験より 20 kHz 以上の音もこのマイクで取得できることを確認したため採用した。マイクをイヤホン部分に固定するため、3D プリンタ (UP Plus2) でマイク取り付け用のアタッチメントを作製し、マイクを固定した (図 3(c))。

なお、利用する周波数領域が 22 kHz までで良い場合は、図 4 に示すように、スマートフォン (Xperia XZ1) とイヤホン、マイクのみで実装可能である。Xperia XZ1 は、ノイズキャンセリングイヤホンのために 5 極ミニプラグに対応しており、ミニプラグでステレオ入力とステレオ出力が可能である。これにより、オーディオインタフェースを省くことで、プロトタイプデバイスを小型化できる。

本研究では、イヤホン、オーディオインタフェース、タブレットを使用してデバイスを実装したが、今後の機器の発展にともない、ヒアラブルデバイス単体で実装可能と考えられる。

4.2 ソフトウェア

提案フレームワークを Android アプリケーション上で実装した (図 3(d))。このアプリケーションでは、3.2 節に示した周波数操作手法をラジオボタンで選択できる。また、各操作手法に対して、スライドバーによって対象とする周波数の詳細な設定が可能である。compression では、ユーザは圧縮後の周波数上限を設定することで、圧縮率を変化できる。overlap では、overlap する領域とされる領域を選択できる。slide では、overlap 同様、slide する領域とされる領域を選択できる。partial elimination/extraction では、削除/抽出する領域を選択できる。この操作により、ユー

ザは自身の要求に応じて、取得する周波数を自在に操作できる。録音と音の発信のサンプリング周波数は 96 kHz とした。FFT のサンプル数は 1,024 とし、512 サンプルずつオーバーラップさせた。

5. 想定アプリケーション

本研究の目的は特定のアプリケーションの提案ではなく、聴力自在化技術のフレームワークの提案であるが、提案フレームワークの適用性を示すために想定アプリケーションを提案する。

5.1 EV/HV の検出

EV や HV は、低速時に静かな走行が可能であるが、そのために歩行者は EV/HV の接近に気づくことができず、事故の原因となっている。一方、これらの車は人には聞こえない高周波のスイッチングノイズを発信していることが知られている [4]。提案フレームワークを用いれば、これらの高周波音を可聴化し、ユーザは EV/HV の接近を知覚できると考えられる。このアプリケーションでは、周波数操作手法として overlap が最適だと考えられる。overlap では、可聴域の音は通常どおりに維持しながら、人に聞こえなかった高域を同時に聞くことができる。そのため、overlap は歩道のない道路脇のような、可聴音も重要かつ背後から接近してくるかもしれない EV/HV の音も同時に取得する必要のある状況で最適な手法だと考えられる。

実際に道路付近で録音した音の周波数スペクトルと環境を図 5 に示す。この図に示すように、25 kHz 付近に特徴的な超音波のピークが確認できる (図 5(b))。20 kHz から 40 kHz を 0 Hz から 20 kHz へと重畳させることで、25 kHz 付近のピークが 5 kHz 付近にも現れる (図 5(c))。これにより、ユーザは EV/HV の高周波音を聞くことができ、これらの接近に気づくことができる。

5.2 異常機器の検出

工場機械などは、異常時に超音波領域の音を発信することや、ガスなどの気体は漏洩時に超音波領域に異音を発することが知られている [5], [6]。このアプリケーションで利用できる周波数操作手法は、compression, overlap, もしくは slide である。compression では、複数周波数の超音波が存在し、これらの位置関係が重要な場合、同時に非可聴域の音を聴取できる。overlap では、可聴音と同時に超音波も聞くことができるため、可聴域が聞こえないと危険な場合に有効である。slide では、対象となる超音波周波数を可聴域に遷移させ、自在に拡大/縮小することで、ユーザにとって最も聞きやすい状態に操作できる。以上より、提案フレームワークを用いれば、これらの異常に自身の聴力を用いて気づき、対応することが可能となる。

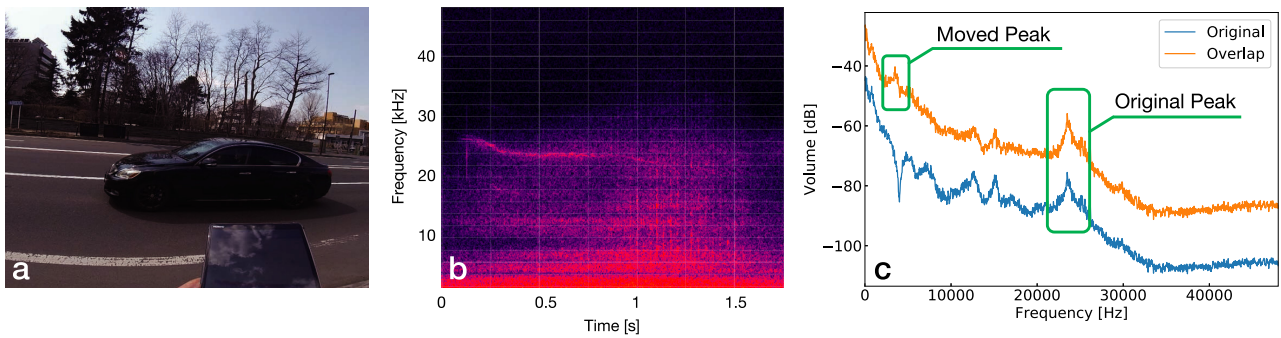


図 5 (a) 音を取得した EV/HV の外観, (b) 対応するスペクトログラム, (c) 操作前後の周波数スペクトル

Fig. 5 (a) Appearance of EV/HV, (b) corresponding spectrogram, and (c) frequency spectrum.

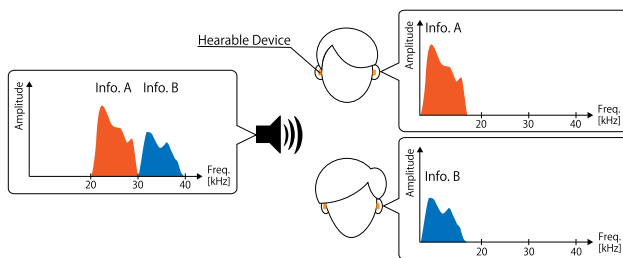


図 6 超音波音声通信

Fig. 6 Ultrasonic voice communication.

5.3 超音波音声通信

提案フレームワークによって、複数の情報を異なる超音波領域に変調させることで、同時に複数の情報を1つのスピーカから提示できる。これは、先行研究 [30] で提案されている手法と同様であるが、提案フレームワークによりウェアラブル環境でも実現が可能と考える。このアプリケーションでは、overlap と slide が適した周波数操作手法だといえる。図 6 に示すように、たとえば、20 k–30 kHz には情報 A を、30 k–40 kHz には情報 B を変調しておく。超音波領域に情報を埋め込んだ音声を超音波スピーカから送信し、周波数操作によってユーザー側で所望の情報の周波数のみを可聴域に変換することで、それぞれのユーザーが異なる情報を得られる。スピーカからの音声を超音波領域であるので、デバイスを用いない人にとっては騒音にはならない。また、Bluetooth のようにデバイス間であらかじめペアリングする必要がない。本手法は、空港や駅など多数の音声情報が存在するが、自分にとって必要のない情報も多く含まれている状況で有用だと考える。

5.4 超音波ヒューマンエコーロケーション

エコーロケーションは、コウモリやイルカなどの動物が使用するソナーである。音波を発信し、周辺の物体からのエコーを利用して、物体の位置や種類を識別している。同様のことが人間に対しても適用でき、一部の視覚障害者は、自分が発する可聴音を用いてエコーロケーションを行うこ

とができる。超音波を利用したエコーロケーションには、物体の高解像度検出や背景音からの抽出の容易さなどの点から利点があるといわれている [16]。提案フレームワークを用いればリアルタイムにエコーを可聴化できるため、超音波ヒューマンエコーロケーションを実現するためのデバイスとして応用できると考える。

5.5 打音検査

打音検査とは、検査用ハンマなどの検査用具により対象物を打撃し、打撃音の変化から変状を判定する診断法である。打音検査は、実施が容易に必要な機器も少ないことから広く用いられている手法である。しかし、打音検査は各ユーザーに一定の経験とスキルが求められる [31]。提案フレームワークを用いれば、重要な周波数に焦点を当てることで、打音検査に熟達していないユーザーをサポートすることが可能である。

実際に取得した打音を図 7 に示す。図 7(a) は、使用したコンクリートとハンマである。コンクリートブロックのサイズは、100 mm × 100 mm × 400 mm であり、1.2 mm × 50 mm の傷が中央につけられている。このコンクリートに対し、正常部と異常部のそれぞれで 3 回打音を記録した。図 7(b) は、正常部と異常部の打撃音の周波数スペクトルを示している。このサンプルでは、正常部でも異常部でも可聴域ではスペクトルが似ており、聞き分けるのは難しいが、30 kHz 付近に周波数の違いがあることが確認できる。したがって、図 7(c) に示すように、slide によって 27 k–31 kHz を 0–4 kHz へと遷移することで、この 2 音の違いを強調できる。

5.6 外界音のイコライジング

提案フレームワークを用いることで、外界音の周波数の一部を強調/抑制したり、抽出/削除したりできる。このアプリケーションには partial elimination/extraction が利用できる。同様の利用方法は Here One [11] でも提案されている。外界音に対して、音楽を自身の好みでイコライズす

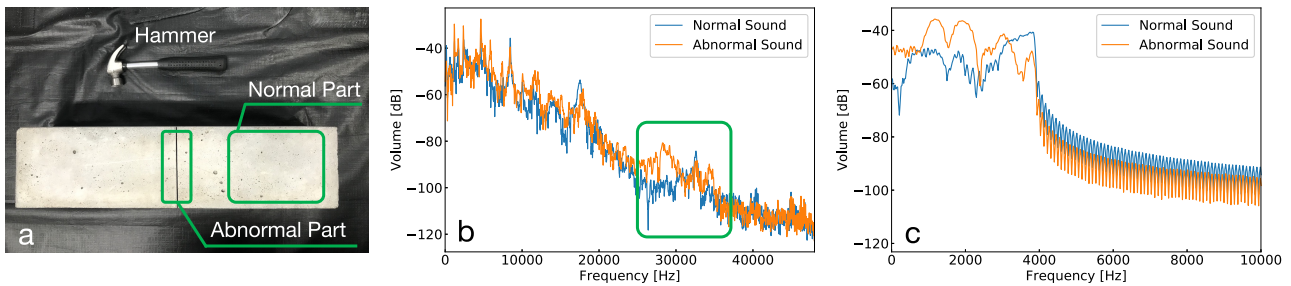


図 7 (a) 使用したコンクリートとハンマ、(b) 操作前の周波数スペクトル、(c) 操作後の周波数スペクトル

Fig. 7 (a) Used concrete and hammer, (b) frequency spectrum before manipulation, and (c) frequency spectrum after manipulation.

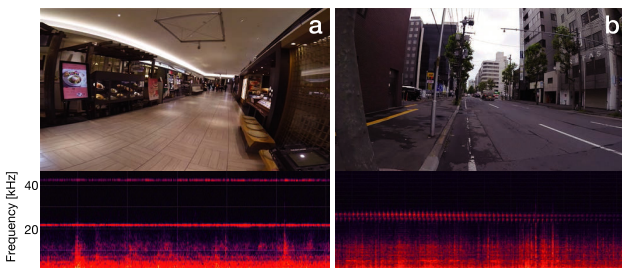


図 8 (a) レストラン付近で確認できた超音波 (b) 道路付近で確認できた超音波

Fig. 8 (a) Observed ultrasound near the restaurant and (b) the road.

るように自由にイコライジングすることが可能である。また、足音の高音を強調することで身体活動のモチベーションが高まると報告されており [13]、音をイコライズした際のユーザへ与える影響の調査には価値があるといえる。提案フレームワークは、ウェアラブル環境で音をリアルタイムに変換でき、今後の外界音を変換する調査に利用できる。

5.7 超音波暴露の検出

公共の場では、車両の検知器や、害獣の撃退器など、様々な用途で超音波が利用されている。これらは大多数の人には聞こえる範囲の音ではないため、人従来の聴力では直接聞こえない。しかし、強力な超音波に長時間暴露されると健康によくないということが報告されている [32]。

実際に街の中で取得できた超音波を図 8 に示す。図 8(a) はレストラン付近で得られた超音波であり、20 kHz と 40 kHz 付近に定常的な強い超音波が確認できる。この超音波は害獣の撃退器の音だと考えられる。図 8(b) は道路で取得した超音波であり、25 kHz 付近にパルス状に短く連続した超音波が確認できる。これは、交通量調査のために利用されている距離センサの超音波だと考えられる。従来の人の聴力ではこれらの音が聞こえないため、超音波領域まで考慮した音空間の設計はなされていない。提案フレームワークでこれらの音を可聴化することで、都市環境の超音波騒音に気づき、街の音空間設計を考え直す機会にすることがで

きる。また、超音波領域も聞こえてしまう一部の人にとっては、partial elimination によって超音波騒音を抑制するという使い方も考えられる。

6. 評価実験

6.1 音源方向の保存

提案フレームワークは実環境で用いることを想定しているが、一部の想定アプリケーション (EV/HV の検出、異常機器の検出、超音波暴露の検出) では、変換後の音源位置を把握することで、ユーザは適切に対応できると考えられる。人間は頭部伝達関数 (HRTF : Head Related Transfer Function) を音源位置の特定に用いているが [33]、提案フレームワークでは周波数を操作するため、音源位置の特定に悪影響を与える可能性がある。したがって、周波数操作後に音源の位置を特定できるかを調査した。

6.1.1 実験方法

実験環境を図 9 に示す。実験は 3 m × 3 m の領域で行った。4 つのスピーカ (Fostex FT200D) を領域の 4 つの角にそれぞれ置いた。これらのスピーカのうち、ランダムでどれか 1 つから音声を発信し、その音源を特定するまでの時間及び被験者の軌跡を記録した。なお、使用するスピーカが偏らないように調整を行った。被験者の軌跡を記録するために、モーションキャプチャシステム (MotionAnalysis MAC3D System) を利用した。被験者は 21–25 歳の男性 8 名である。利用した音源は 6 種類であり、440 Hz の純音 (S1)、0–20 kHz の帯域制限ノイズ (S2)、0–20 kHz 帯域の可聴音のアナウンス (S3)、20,440 Hz の純音 (S4)、20 k–40 kHz の帯域制限ノイズ (S5)、20 k–40 kHz 帯域の超音波アナウンス (S6) である。純音は単一の周波数しか含まないため、比較のために指定帯域の全周波数を含む帯域制限ノイズおよび想定アプリケーション (超音波音声通信) で用いる予定のアナウンス音を用いて実験を行った。被験者が超音波 (S4, S5, S6) を聞く際には、slide によって可聴音に変換した音声を聞いた。具体的には、S4 は 440 Hz に、S5 と S6 は 0–20 kHz へと変換された。被験者に提示する音圧は、事前に slide によって可聴化された音を聞く

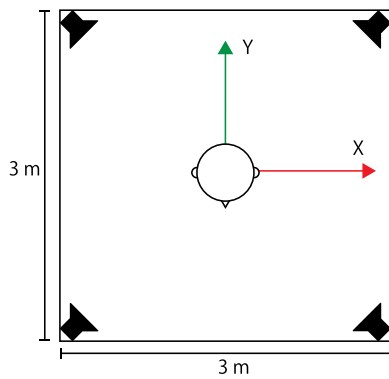


図 9 実験環境

Fig. 9 Environment of the experiment.

表 1 音源特定までにかかった時間 [s]

Table 1 Time for sound source detection [s].

Participant	Sound Source					
	S1	S2	S3	S4	S5	S6
A	3.5	3.5	2.7	3.2	6.6	1.8
B	8.7	3.0	3.1	13.8	5.8	6.1
C	5.0	1.8	5.2	9.5	3.7	6.6
D	5.9	3.7	2.4	8.5	5.6	5.5
E	5.0	6.7	13.4	8.1	5.0	2.7
F	5.6	5.1	4.6	8.6	6.3	9.5
G	1.4	1.3	2.9	7.4	6.4	18.1
H	5.2	3.1	6.2	5.0	9.6	3.9
Average	5.0	3.5	5.1	8.0	6.1	6.8

ことで、ふだんの自身の耳の聞こえ方と同程度と感ずるよう被験者自身で設定してもらった。ここでの想定アプリケーションでは、overlap と slide が適した周波数操作手法といえる。本調査では、操作後の音の影響のみについて着目したかったため、操作した音以外の影響が少ない slide を周波数操作手法として選択した。

6.1.2 結果

実験の結果を表 1 に示す。この表に示すように、音源特定までの時間は、S1 では 5s, S2 では 3.5s, S3 では 5.1s, S4 では 8.0s, S5 では 6.1s, S6 では 6.8s となった。また、すべての被験者は正しい音源を特定することができた。図 10 に、周波数操作前と後で音源特定までにかかった時間を比較したものを示す。図中のエラーバーは標準偏差を示す。すべての音源において、周波数操作後の方が音源特定に時間がかかっている。しかし、音源を対象とした 1 要因分散分析を行った結果、6 つの音源特定時間の間に有意差は確認できなかった。

ここでの想定アプリケーションでは、危険性の観点から、EV/HV の接近検知が最も音源方向特定の性能を求められるアプリケーションだと考えられる。EV/HV から発信される超音波に最も近い音は、この調査では S4 にあたる。図 11 に、音源特定までに最も時間のかかった被験者 B の S4 のときの軌跡を示す。原点と x 軸, y 軸は図 9 に対応し

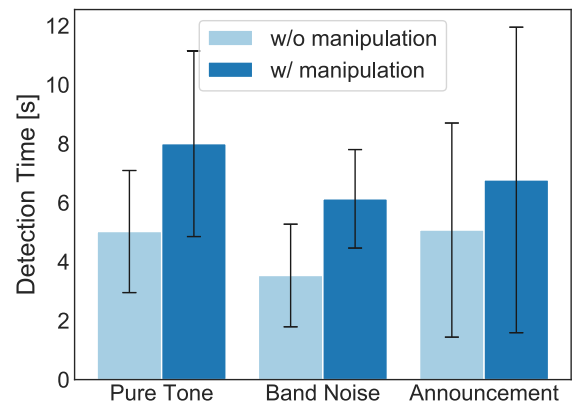


図 10 周波数操作無し/ありの結果の比較

Fig. 10 Comparison of with/without manipulation. The bars show standard deviations.

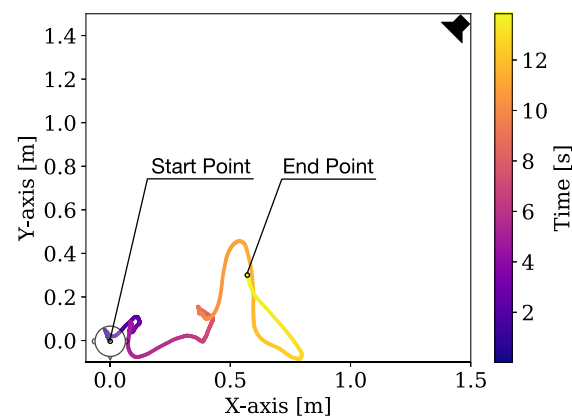


図 11 被験者 B の S4 の時の軌跡

Fig. 11 Trajectory of participant B for S4.

ており、右上のスピーカが音源の位置であった。この図が示すように、被験者は約 2 秒の段階で音源の大まかな場所を把握し、そちらに移動を初めている。しかし、正解に確信を得るために、他のスピーカとの間を行き来することで音源特定までに時間がかかっていた。また、すべての被験者が、「左右の耳の音量差を利用して音源方向を特定できた」とコメントしていた。提案フレームワークでは、周波数操作によって HRTF が変化し、周波数操作前よりも音源方向の特定が難しくなるといえる。しかし、両耳間レベル差 (ILD: Interaural Level Difference) は維持され、被験者は ILD を頼りに大まかな音源方向を把握できる。実験結果より、被験者は操作後の音でも早い段階でおおまかな音源方向を特定できることが示唆された。ユーザが超音波を感知し、そちらの方向を向くと、ユーザは EV/HV の存在に気づくことができる。したがって、想定アプリケーションでの利用を考えると、周波数操作後の音源方向の保存は提案フレームワークの性能で問題ないと考えられる。

6.2 周波数操作後の遅延

提案フレームワークでは、ソフトウェア上で周波数操作を行っているため、変換による遅延が生じる。ユーザの視

表 2 各 FFT サイズにおける計算時間 [ms].

Table 2 Computation time of each FFT size [ms].

Size	Normal	Compression	Overlap	Slide	PEL ¹	PEX ²
512	0.005	0.015	0.012	0.018	0.016	0.015
1,024	0.009	0.031	0.023	0.034	0.031	0.029
2,048	0.010	0.053	0.036	0.059	0.053	0.050
4,096	0.011	0.109	0.072	0.123	0.110	0.102

¹Partial Elimination ²Partial Extraction

表 3 各 FFT サイズにおけるシステム全体の遅延 [ms]

Table 3 Overall system delay of each FFT size [ms].

FFT Size	512	1,024	2,048	4,096
Delay	136.3	140	144.9	156.9

覚から得られる情報と、変換された音との間に大幅な遅延があると、ユーザにとって不快となる可能性がある [34], [35]. 変換による音の遅延の調査は補聴器の分野でも行われているが [36], 本研究では周波数操作を行っている. そこで本調査では, 周波数操作により生じる遅延がユーザに与える影響を調査した. 提案フレームワークでは, 遅延は主に FFT サイズによって生じる. FFT サイズが, データの取得/発信時間や計算時間に影響を与える. そのため, 本調査では 4 つの FFT サイズ (512, 1,024, 2,048, 4,096) について調査した.

6.2.1 遅延時間の測定

まず, 各 FFT サイズにおいて, 周波数操作手法の違いによってどの程度の遅延が生じるのかを調査した. 10,000 回の計算時間の平均値を表 2 に示す. ここで, normal は周波数スペクトル上で何も操作を行わないことを示す. 周波数操作手法による計算時間の差は, 最大でも FFT サイズ 4,096 のときの normal と slide の間で 0.11 ms であった. これは, 人間の知覚できる遅延よりも大幅に小さいため, 周波数操作手法による人への遅延の影響は大きくないと考える [37].

次に, 実際にユーザが体感する遅延, すなわちデバイスに音が入力されてから周波数操作後の音が出力されるまでの時間を計測した. 表 2 より周波数操作の有無で大きな時間差は確認できなかったため, 周波数操作手法は normal とした. 100 回の計測結果の平均を表 3 に示す. この表に示すように, FFT サイズが最小の 512 でも 136.3ms の遅延があり, FFT サイズが大きくなるほど遅延時間が大きくなることが確認できる.

6.2.2 ユーザ評価

各 FFT サイズで生じる遅延によって, ユーザが受ける印象を調査した. 被験者は, 実装したプロトタイプデバイスを用いて変換後の音声を聞く. 上記の結果で周波数操作手法による大きな差はないことが確認できたため, 周波数操作手法として overlap を選択した. 被験者には, デバイスを装着して変換後の音声を聴きながら日常的な動作 (タ

表 4 遅延実験でのスコアと対応する評価

Table 4 Defined score and corresponding delay.

Score	Delay
5	遅延がまったく認められない
4	遅延が認められるが, 気にならない
3	遅延がわずかに気になる
2	遅延が気になる
1	遅延が非常に気になる

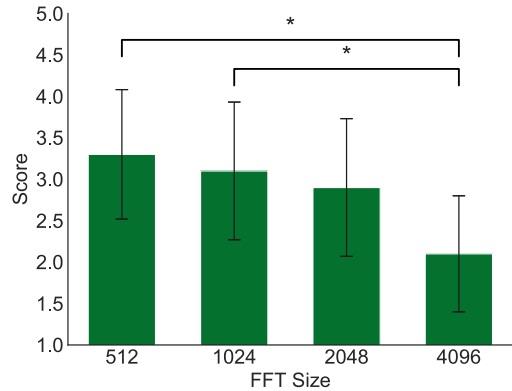


図 12 各 FFT サイズにおける遅延の結果

Fig. 12 Result of delay. The bars show standard deviations.

イピング, 動画の視聴, 筆記, 物を叩くなど) を行ってもらう. 被験者は, 自分の行った動作や視覚から得られた情報と, デバイスから聞こえてくる音の遅延についてどう感じるか評価を行う. 被験者に提示する音圧は, 机を叩くなどによって得られる音を聞くことで, ふだんの自身の耳の聞こえ方と同程度と感ずるように被験者自身で設定してもらった. 評価指標は 1-5 とし, デバイスを装着していない状態の普段の自分自身の耳と比較して, 表 4 に示す尺度で評価を行った. 被験者は 22-29 歳の男性 9 名と女性 1 名であり, 調査した FFT サイズの順番はランダムとした.

実験結果を図 12 に示す. 図中のエラーバーは標準偏差を示す. スコアは, FFT サイズが 512 で 3.3, 1,024 で 3.1, 2,048 で 2.9, 4,096 で 2.1 となった. 分散分析の結果, FFT サイズによる主効果が有意であった ($F(3, 27) = 6.12, p < 0.01, f = 0.82$). Bonferroni 法による多重比較を行った結果, 512 と 4,096, 1,024 と 4,096 の間に有意差が確認できた ($p < 0.05$). 以上より, 遅延に関しては, FFT サイズは小さいほどよく, 大きいほど不快となることが確認できた.

最も高いスコアでも, FFT サイズが 512 のときの 3.3 であり, 十分に高い値とはいえない. これは, 表 3 に示すように, FFT サイズが 512 でもシステム全体として 136.3ms の遅延が発生しているためと考えられる. また, 被験者 3 からは「FFT サイズが 4,096 のとき, 最初は不快感を感じたが, 慣れると不快感が少なくなった」というコメントが得られた. したがって, 長時間装着することで, ユーザは遅延に適応できる可能性がある. 長時間装着による影響に

表 5 音質実験でのスコアと対応する評価

Table 5 Defined score and corresponding impairment.

Score	Impairment
5	劣化がまったく認められない
4	劣化が認められるが、気にならない
3	劣化がわずかに気になる
2	劣化が気になる
1	劣化が非常に気になる

については 7.8 節で議論を行う。

6.3 周波数操作後の音質

3 章で述べたように、フェーズボコーダでは、音質の劣化を完全になくすことが難しい [29]。アプリケーションを想定すると、最も音質が要求されるのは、超音波音声通信だと考える。他の想定アプリケーション (EV/HV の検出、異常機器の検出) では、ふだん聞こえない音が聞こえるようになるだけでも有用であるが、超音波音声通信では音声の内容まで聞き取って理解する必要があるため、周波数操作後の音質が重要と考えた。したがって、本節では周波数操作後の音質を調査した。

6.3.1 実験方法

超音波音声通信を想定して、復調後の音質と、音声の意味が理解できるか調査を行った。利用した音声は駅でのアナウンスと空港でのアナウンスを想定したものである。駅でのアナウンスを 21 kHz から 29 kHz に、空港でのアナウンスを 30 kHz から 38 kHz へ変調して、スピーカ (Fostex FT200D) から発信する。これらを、実装したデバイスで slide によって 0 Hz から 8 kHz へと遷移することで、可聴音に復調する。被験者に提示する音圧は、事前に音圧設定用のサンプル音声を聞くことで、普段の自身の耳の聞こえ方と同程度と感ずるよう被験者自身で設定してもらった。周波数分解能は FFT サイズによって変化する。そのため、512, 1,024, 2,048, 4,096 の 4 通りの FFT サイズについて音質を調査した。音源はすべて被験者の母国語 (日本語) で作成した。評価指標として、Degradation Mean Opinion Score (DMOS) を用いた [38]。被験者は、変調前の音源を聞いた直後に変調後の音源を聞くことで、どれだけ音源が劣化したかを 1-5 の尺度で評価を行う。評価は、表 5 に示す尺度に基づいて行われた。被験者は、6.2 節と同様の 22-29 歳の男性 9 名と女性 1 名である。

6.3.2 結果

図 13 と図 14 に結果を示す。図中のエラーバーは標準偏差を示す。21 k-29 kHz で変調した場合、FFT サイズが 512 では 2.2, 1,024 では 3.8, 2,048 では 3.4, 4,096 では 4.1 となった。30 k-38 kHz で変調した場合、FFT サイズが 512 では 1.8, 1,024 では 2.9, 2,048 では 2.6, 4,096 では 3.5 となった。分散分析の結果、20 kHz 台を用いた際

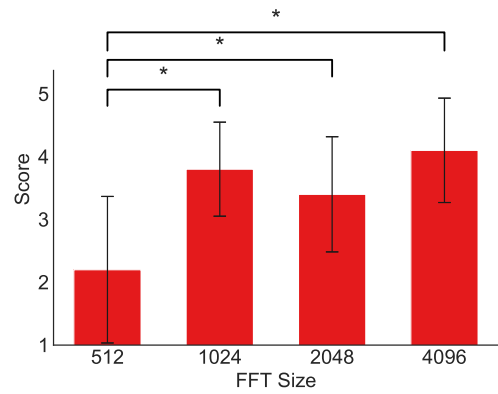


図 13 20 kHz 台を用いた時の各 FFT サイズにおける音質の結果
Fig. 13 Result of sound quality using 20 kHz range. The bars show standard deviations.

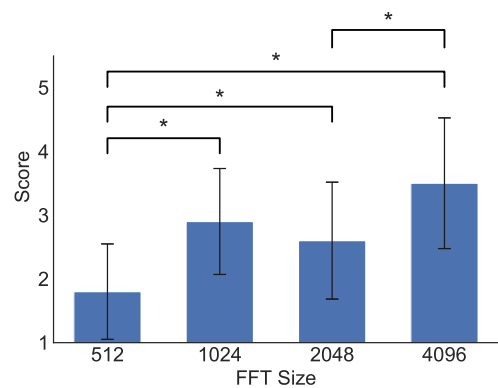


図 14 30 kHz 台を用いた時の各 FFT サイズにおける音質の結果
Fig. 14 Result of sound quality using 30 kHz range. The bars show standard deviations.

($F(3, 27) = 12.22, p < 0.01, f = 1.17$), 30 kHz 台を用いた際 ($F(3, 27) = 16.87, p < 0.01, f = 1.37$) とともに FFT サイズによる主効果が有意であった。Bonferroni 法による多重比較の結果、20 kHz 台の場合、512 と 1,024 の間、512 と 2,048 の間、512 と 4,096 の間に、それぞれ有意差が確認できた ($p < 0.05$)。また、30 kHz 台の場合、512 と 1,024 の間、512 と 2,048 の間、512 と 4,096 の間、2,048 と 4,096 の間に、それぞれ有意差が確認できた ($p < 0.05$)。

結果が示すように、FFT サイズが 512 の時が最もスコアが低く、FFT サイズが 4,096 のときに最もスコアが高くなった。21 k-29 kHz と 30 k-38 kHz を用いた際の結果を比較すると、21 k-29 kHz を用いた結果の方が良いスコアが得られていた。これは、高周波の方が空气中で減衰しやすいことと、マイクの周波数特性が高周波に行くほど悪かったためと考えられる。しかし、全被験者がすべての変調後の音源の意味が聞き取れたと回答しており、内容の理解には問題がないと考えられる。したがって、音声劣化は確認できるが、提案フレームワークによる超音波音声通信アプリケーションは実現可能だと考えられる。

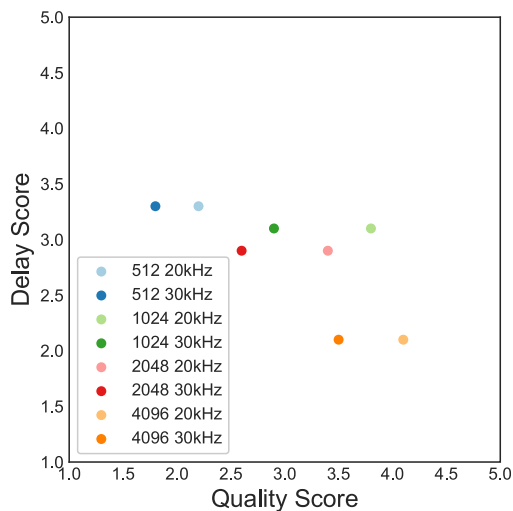


図 15 遅延と音質の結果
Fig. 15 Result of sound delay vs. quality.

7. 考察

7.1 遅延と音質のトレードオフ

3章で述べたように、FFTを利用するには時間分解能と周波数分解能はトレードオフの関係にある。図15に音の遅延と音質の結果を示す。この図において、右上にある点ほど、良い性能を示すといえる。20kHz台と30kHz台の両方において、FFTサイズ1,024が最も右上に近い点に位置していることが確認できる。したがって、本研究においては、音の遅延と音質を考慮するとFFTサイズ1,024が最適だといえる。

7.2 直接音の影響

遅延による不快感の原因の1つが、直接音によるものである。本研究では、高遮音性のイヤホンを利用したが、特に大きな音の場合は完全に外界からの直接音を遮断することは難しい。そのため、ユーザは直接音の直後に変換後の音を聞くことになり、エコー効果のような聞こえ方になってしまう場合がある。この不快感を減少させるため、外耳道を覆わないイヤホンを用いることが考えられる。これにより、可聴音はユーザ従来の聴力で聴きながら非可聴音のみをデバイスで変換して聞くことができる。非可聴音はユーザ本来の聴力では取得できないため、遅延による不快感は発生しないと考える。この実装はoverlapの時のみ有効であるが、overlapのみが必要な場面では効果的だと考える。

7.3 デバイス性能

本研究では96kHzのサンプリングを行ったが、一般的なヒアラブルデバイスはまだ96kHzには対応していない。また、オーディオインタフェースを経由しても96kHzで録音できるデバイスは、今回テストした中ではMediapad

M5だけであった。現状ではまだ市販のヒアラブルデバイスのみで提案フレームワークを実装することは難しいが、本論文で述べたように、市販のモバイル/ウェアラブルデバイスを組み合わせることで、提案フレームワークは実現可能である。今後の機器の発展にともない、ヒアラブルデバイスのみでの提案フレームワークの実装が可能になると考える。

7.4 機械による検出と人による検出

想定アプリケーションとして、EV/HVの検出や異常機器の検出、打音検査などをあげたが、これらは機械学習を用いて自動検知することも考えられる。しかし、機械学習ベースの手法だと学習していないコンテキストは検出することが難しい。我々は、機械に状況の判断をすべて任せるのではなく、機械が人間の能力を拡張し、補助することによって、状況の判断は人間自身が行うことが重要と考え、本手法を提案した。聴力自在化技術による手法と機械学習による手法は、どちらがどのような環境で有効であるか、今後調査を行う必要がある。

7.5 マイクの方向

本論文では、実装のしやすさを考慮し、マイクの方向をユーザの正面方向とした。しかし、実際の人間の耳を考えると、外耳道は頭部横方向についており、耳介が外耳道を覆うようになっている。人がより自然に感じるマイクの方向や、外耳道型のカバーを装着するべきかなどの調査は今後の課題である。

7.6 時間と空間への適用

本研究では、音の周波数の操作に着目したが、提案フレームワークは時間や空間方向への応用も考えられる。時間の操作としては、取得した音の再生速度を操作することで、聞き逃した音をもう1度聞き返すことや、早口な人の音声を聞きやすくするなどの操作が可能だと考える。また、未来の音を予測して聞くということも考えられる。今後起こりうる未来の音を聞くことができれば、危険の察知など、今後起こりうる事象への対処が可能になる。

空間の操作としては、特定方向の音量を増加/減少させることや、ある音源からの音を本来とは異なる場所から聞こえるように感じさせるなど、音の空間に関わる情報を操作することが考えられる。この操作により、ユーザの集中したい方向だけの音の取得や、音源位置操作によるナビゲーションなどの応用が考えられる。

さらに、周波数、時間、空間の操作を組み合わせることで、より自由自在な聴力の操作が可能になると考えられる。

7.7 想定アプリケーションの実装

本研究では、聴力自在化技術のフレームワークの提案に

着目したため、各アプリケーション実装の際の問題点については考慮していない。たとえば、EV/HVの検出や打音検査技術の向上においては、これらのアプリケーションに効果的な周波数がどこに存在するかをユーザが効率的に知る方法や、それをどこに可聴化するのが最適であるかなどは考慮されていない。超音波音声通信においても、自分の欲しい情報がどの帯域に存在しているのかをユーザが知る方法は考慮されていない。また、これらのアプリケーションについて定量的な評価はされていない。さらに、聴力には個人差があると考えられるため、個人による効果の差についても調査を行う必要がある。各アプリケーションの実装、考慮すべき問題の解決や評価は今後の課題である。

7.8 長期装着の影響

6.2節の遅延の調査において、最初は遅延が気になったが、装着が長くなるにつれ遅延が気にならなくなったという意見が被験者から得られた。視覚の分野では、プリズムによって視野を平行にずらすと、ターゲットに向かって手を伸ばす運動に初めは誤差を生じるが、試行とともに回復するプリズム適応という現象が確認されている [39]。聴覚においても同様に、音の遅延に対して適応できることが報告されており、システムを長期間利用することで遅延に適応できる可能性がある [37], [40], [41]。長期装着による操作後の音への適応については今後調査を行う必要がある。

また、長時間利用による健康上の問題についても考慮する必要がある。WHOによると、個人用オーディオ機器を用いた際の安全なリスニングは、音圧レベルおよびリスニングの継続時間と頻度に依存するとされている [42], [43]。本研究では音圧レベルはユーザ側で自在に設定できるが、音圧レベルの上限を設定することや、一定以上の音圧レベルになるときはシステム側で警告を行うなど、過度な音圧レベルをユーザに提示することがないような対応が考えられる。継続時間と頻度に関しても、システム側で音圧レベルと使用時間を計測し、人の許容量を超えることのないようにモニタリングすることで対応できると考える。

一方、本研究で提案する周波数操作を長期間適用した際の健康への影響は、我々の知る限り調査されていないため、今後調査を行う必要がある。

8. おわりに

本研究では、外界音の周波数操作による聴力の自在化技術を提案した。5つの周波数操作手法を提案し、プロトタイプデバイスを実装した。また、提案フレームワークによって実現可能な7つの想定アプリケーションを提案した。実装したデバイスを用いた調査の結果、周波数操作後においても音源位置の検出が可能なが確認できた。また、周波数操作にともなう音の遅延と音質のトレードオフについても議論した。提案フレームワークは、今後ウェアラブル

コンピューティング環境において音を変換する研究に適用できると考える。

謝辞 本研究はJSPS科研費JP18K18084とMSRA CORE14の助成を受けたものである。

参考文献

- [1] Langlotz, T., Sutton, J., Zollmann, S., Itoh, Y. and Regenbrecht, H.: ChromaGlasses: Computational Glasses for Compensating Colour Blindness, *Proc. 2018 CHI Conference on Human Factors in Computing Systems, CHI'18*, pp.390:1–390:12 (2018).
- [2] Ardouin, J., Lécuyer, A., Marchal, M., Riant, C. and Marchand, E.: FlyVIZ: A Novel Display Device to Provide Humans with 360 Vision by Coupling Catadioptric Camera with Hmd, *Proc. 18th ACM Symposium on Virtual Reality Software and Technology, VRST '12*, pp.41–44 (2012).
- [3] Ates, H.C., Fiannaca, A. and Folmer, E.: Immersive Simulation of Visual Impairments Using a Wearable See-through Display, *Proc. 9th International Conference on Tangible, Embedded, and Embodied Interaction, TEI '15*, pp.225–228 (2015).
- [4] Takagi, M., Fujimoto, K., Kawahara, Y. and Asami, T.: Detecting Hybrid and Electric Vehicles Using a Smartphone, *Proc. 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '14*, pp.267–275 (2014).
- [5] 3DSignals: Sound-based Predictive Maintenance (2019), available from (<https://www.3dsig.com/>).
- [6] Emerson: About Ultrasonic Gas Leak Detectors (2019), available from (<https://www.emerson.com/en-sg/automation/measurement-instrumentation/flame-gas-detection/about-ultrasonic-gas-leak-detectors.>)
- [7] Watanabe, H. and Terada, T.: Manipulatable Auditory Perception in Wearable Computing, *Proc. Augmented Humans International Conference, AHs '20*, pp.9:1–9:12, ACM (online), DOI: 10.1145/3384657.3384792 (2020).
- [8] Pettersson Elektronik: Bat detectors and sound analysis software (2019), available from (<http://www.batsound.com/>).
- [9] ReSound: Sound Shaper - Newest hearing aids tailored to your individual hearing environment and preference (2019), available from (<http://www.gnhearing.dk/en/professionals/technology-and-innovation/surround-sound-by-resound/sound-shaper>).
- [10] Oticon: Inium Sense - Speech Rescue™ technology (2019), available from (<https://www.oticon.com/professionals/brainhearing-technology/brainhearing-technology/speech-rescue>).
- [11] Doppler Labs: Here One (2019), available from (<https://hereplus.me/products/here-one/>).
- [12] Bose: Bose Headphones (2019), available from (https://www.bose.com/en_us/products/wellness/conversation_enhancing_headphones/headphones.html).
- [13] Tajadura-Jiménez, A., Basia, M., Deroy, O., Fairhurst, M., Marquardt, N. and Bianchi-Berthouze, N.: As Light As Your Footsteps: Altering Walking Sounds to Change Perceived Body Weight, Emotional State and Gait, *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pp.2943–2952 (2015).
- [14] Tajadura-Jiménez, A., Newbold, J., Zhang, L., Rick, P. and Bianchi-Berthouze, N.: As Light As You Aspire to

- Be: Changing Body Perception with Sound to Support Physical Activity, *Proc. 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, pp.658:1–658:14, ACM (online), DOI: 10.1145/3290605.3300888 (2019).
- [15] Uchibori, S., Sarumaru, Y., Ashihara, K., Ohta, T. and Hiryu, S.: Experimental evaluation of binaural recording system using a miniature dummy head, *Acoustical Science and Technology*, Vol.36, No.1, pp.42–45 (online), DOI: 10.1250/ast.36.42 (2015).
- [16] Sumiya, M., Ashihara, K., Yoshino, K., Gogami, M., Nagatani, Y., Kobayasi, K.I., Watanabe, Y. and Hiryu, S.: Bat-inspired signal design for target discrimination in human echolocation, *The Journal of the Acoustical Society of America*, Vol.145, No.4, pp.2221–2236 (2019).
- [17] Sohl-Dickstein, J., Teng, S., Gaub, B.M., Rodgers, C.C., Li, C., DeWeese, M.R. and Harper, N.S.: A device for human ultrasonic echolocation, *IEEE Trans. Biomedical Engineering*, Vol.62, No.6, pp.1526–1534 (2015).
- [18] Weger, M., Hermann, T. and Höldrich, R.: Real-time auditory contrast enhancement, *International Conference on Auditory Display, ICAD '19*, pp.254–261, Georgia Institute of Technology (2019).
- [19] Albrecht, R., Lokki, T. and Savioja, L.: A Mobile Augmented Reality Audio System with Binaural Microphones, *Proc. Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications, IwS '11*, pp.7–11, Association for Computing Machinery (online), DOI: 10.1145/2019335.2019337 (2011).
- [20] Storek, D., Stuchlik, J. and Rund, F.: Modifications of the surrounding auditory space by augmented reality audio: Introduction to warped acoustic reality, *International Conference on Auditory Display, ICAD '15*, pp.225–230 (2015).
- [21] Bedri, A., Byrd, D., Presti, P., Sahni, H., Gue, Z. and Starner, T.: Stick It in Your Ear: Building an In-ear Jaw Movement Sensor, *Adjunct Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct*, pp.1333–1338 (2015).
- [22] Bedri, A., Verlekar, A., Thomaz, E., Avva, V. and Starner, T.: Detecting Mastication: A Wearable Approach, *Proc. 2015 ACM on International Conference on Multimodal Interaction, ICMI '15*, pp.247–250, ACM (online), DOI: 10.1145/2818346.2820767 (2015).
- [23] Ando, T., Kubo, Y., Shizuki, B. and Takahashi, S.: CanalSense: Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals, *Proc. 30th Annual ACM Symposium on User Interface Software and Technology, UIST '17*, pp.679–689 (2017).
- [24] Manabe, H., Fukumoto, M. and Yagi, T.: Conductive Rubber Electrodes for Earphone-based Eye Gesture Input Interface, *Proc. 2013 International Symposium on Wearable Computers, ISWC '13*, pp.33–40, ACM (online), DOI: 10.1145/2493988.2494329 (2013).
- [25] Laput, G., Chen, X.A. and Harrison, C.: SweepSense: Ad Hoc Configuration Sensing Using Reflected Swept-Frequency Ultrasonics, *Proc. 21st International Conference on Intelligent User Interfaces, IUI '16*, pp.332–335, ACM (online), DOI: 10.1145/2856767.2856812 (2016).
- [26] Arakawa, T., Koshinaka, T., Yano, S., Irisawa, H., Miyahara, R. and Imaoka, H.: Fast and accurate personal authentication using ear acoustics, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA '16*, pp.1–4 (2016).
- [27] NEC: Human-Oriented IoT Solutions Using Hearable Technology from NEC (2019), available from (<https://www.nec.com/en/global/techrep/journal/g17/n01/170110.html>).
- [28] Amesaka, T., Watanabe, H. and Sugimoto, M.: Facial Expression Recognition Using Ear Canal Transfer Function, *Proc. 23rd International Symposium on Wearable Computers, ISWC '19*, pp.1–9 (2019).
- [29] Laroche, J. and Dolson, M.: New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects, *Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA '99*, pp.91–94 (1999).
- [30] 吉野祥之, 苗村 健: u-soul: 超音波を用いた空間的な画像提示システム, 日本バーチャルリアリティ学会論文誌, Vol.13, No.2, pp.239–246 (2008).
- [31] Yamashita, A., Hara, T. and Kaneko, T.: Inspection of Visible and Invisible Features of Objects with Image and Sound Signal Processing, *Proc. 2006 IEEE International Conference on Intelligent Robots and Systems*, pp.3837–3842 (2006).
- [32] Fletcher, M.D., Lloyd Jones, S., White, P.R., Dolder, C.N., Lineton, B. and Leighton, T.G.: Public exposure to ultrasound and very high-frequency sound in air, *The Journal of the Acoustical Society of America*, Vol.144, No.4, pp.2554–2564 (2018).
- [33] Blauert, J.: *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press (1996).
- [34] 樋田浩一, 浅井康平, 上野佳奈子, 鶴田総太郎: 視聴覚と運動の非同時性判断に関する研究, 日本認知科学会第28回大会発表論文集, pp.482–485 (2011).
- [35] 栗田孝昭: 音声遅延と映像遅延が通信品質に及ぼす影響, 日本音響学会誌, Vol.52, No.1, pp.57–62 (オンライン), DOI: 10.20697/jasj.52.1.57 (1995).
- [36] Alexander, J.: Hearing Aid Delay and Current Drain in Modern Digital Devices, *Canadian Audiologist*, Vol.3, No.4 (2016).
- [37] 樋田浩一, 上野佳奈子, 嶋田総太郎: 身体運動に伴う遅延聴覚フィードバックの知覚順応, 認知科学, Vol.20, No.4, pp.493–497 (オンライン), DOI: 10.11225/jcss.20.493 (2013).
- [38] International Telecommunication Union: P.800: Methods for subjective determination of transmission quality (2019), available from (<https://www.itu.int/rec/T-REC-P.800-199608-I/en>).
- [39] Facchin, A., Folegatti, A., Rossetti, Y. and Farné, A.: The half of the story we did not know about prism adaptation, *Cortex*, Vol.119, pp.141–157 (online), DOI: 10.1016/j.cortex.2019.04.012 (2019).
- [40] 藤崎和香, 西田眞也: 異種感覚モダリティ間の同期知覚, 基礎心理学研究, Vol.27, No.2, pp.160–163 (オンライン), doi10.14947/psychono.KJ00005578159 (2009).
- [41] Vatakis, A., Navarra, J., Soto-Faraco, S. and Spence, C.: Temporal recalibration during asynchronous audiovisual speech perception, *Experimental Brain Research*, Vol.181, No.1, pp.173–181 (2007).
- [42] World Health Organization: 1.1 billion people at risk of hearing loss (2015), available from (<https://www.who.int/mediacentre/news/releases/2015/ear-care/en/>).
- [43] World Health Organization and International Telecommunication Union: Safe Listening Devices and Systems: A WHO-ITU Standard (2019), available from (<https://www.who.int/publications/m/item/safe-listening-devices-and-systems-a-who-itu-standard>).

apps.who.int/iris/bitstream/handle/10665/280085/9789241515276-eng.pdf).

推薦文

本論文は、ユーザの耳元のマイクで取得した外界音の周波数を操作し、変換した音をリアルタイムでユーザに提示するフレームワークを、「聴力自在化」という発展性のあるコンセプトとともに提案している。コンセプトを具現化する7種類のアプリケーションも説得力が高く、評価実験も様々な角度から実施され、音源方向、遅延、音質に関する知見が考察されていて学術的にも価値が高い。論文も分かりやすく記述されており完成度はきわめて高い。総じてプログラム委員会において高い評価を受けたため、推薦論文として推薦する。

(インタラクション 2020 プログラム委員長 志築 文太郎)



渡邊 拓貴 (正会員)

2012年神戸大学工学部電気電子工学科卒業。2014年同大学大学院工学研究科博士前期課程修了。2017年同大学大学院工学研究科博士後期課程修了。同年より北海道大学大学院情報科学研究科助教、現在に至る。博士(工学)。ウェアラブルコンピューティング、ユビキタスコンピューティングの研究に従事。ACM会員。



寺田 努 (正会員)

1997年大阪大学工学部情報システム工学科卒業。1999年同大学大学院工学研究科博士前期課程修了。2000年同大学院工学研究科博士後期課程退学。同年大阪大学サイバーメディアセンター助手。2005年同講師。2007年神戸大学大学院工学研究科准教授。2018年同大学院工学研究科教授、現在に至る。博士(工学)。ウェアラブルコンピューティング、ユビキタスコンピューティングの研究に従事。ACM, IEEE, ヒューマンインタフェース学会等, 7学会の会員。