

ウェアラブルコンピューティングにおける 聴力自在化技術の提案

渡邊 拓貴^{1,a)} 寺田 努^{2,b)}

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（ヒアラブルデバイス）に着目した。ヒアラブルコンピューティング環境では、ユーザは常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。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聴取や、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。

1. はじめに

近年のコンピュータの小型化に伴い、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、デバイスを常に装着しているため、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。例えば、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。特に近年着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、通話等も可能である。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、ユーザは常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。一方、人間の耳は自らの意志では制御が難しく、取得したい音の取舍選択をすることは困難である。例えば、ユーザが自身のタスクに集中したい時に他人の喋り声や騒音など、外界音に集中を阻害される状況が考えられる。また、人間が聞くことができる音はおよそ 20Hz から 20kHz といわれており、多くの環境音は 20kHz 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、20kHz 以上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なければ、思わぬ事故に繋がる可能性がある。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なが確認できた。また、周波数操作に伴う音の遅延と音質のトレードオフについても調査を行なった。

¹ 北海道大学

² 神戸大学

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

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

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 [7]。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 [8]。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 [9]。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 [10]。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 [11]。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 [12]。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。Sumiya らは、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）の可能性を調査している [13]。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覚自在化のためのフレームワークを提供しており、これらの関連研究で行われていることは、本研究に含ま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、本研究のシステム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されている。Bedri らは、赤外線近接センサを用いて下顎の動きに伴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 [14]。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 [15]。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 [16]。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 [17]。Amesaka らは、外耳道の伝達関数を利用することでユーザの顔の動作を認識している [18]。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。提案手法では、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聴取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 [19]。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等の手法 [7] 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。提案手法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、フェーズボコーダは音声劣化を完全に無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 [19]。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 節で述べられ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、ユーザはスピーカから発信される操作された音を聴取する。ユーザは、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 96kHz とした。FFT のサンプル数は 1024 とし、サンプル数の半分のスライディング幅とした。FFT を計算する際には、周波

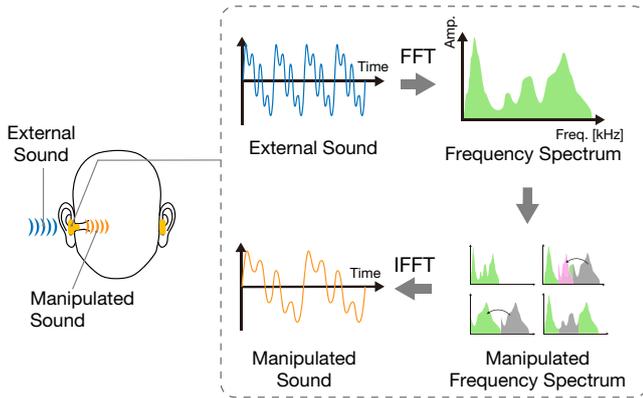


図 1 システム構成

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 1024 とした。詳細な議論は 6.4 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、ユーザは聴力を自在に操作できる。提案手法の重要な特徴の一つは、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 2a 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時に利用できる。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 2b)。本手法は Speech Rescue[9] で用いられているものに近いが、提案手法では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聴取できる。したがって、本手法は混雑した道や道路のような、従来の可聴音が無くなると危険な場面での利用を想定している。

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 2c)。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。例えば、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

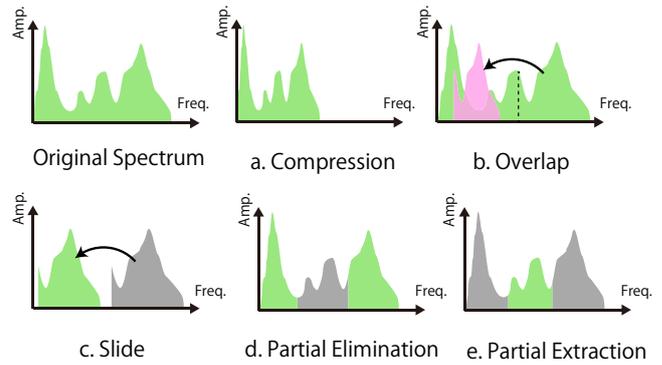


図 2 周波数操作手法: (a) compression (b) overlap (c) slide (d) partial elimination (e) partial extraction

取る場合にも利用できる。

3.2.4 Partial Elimination

partial elimination は、図 2d 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 2e 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。筆者ら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 96kHz 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。可能な限り外界音を遮断し、変換後の音のみを取得するようにするため、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 30Hz から 20kHz となっているが、予備実験より 20kHz 以上の音でも取得できることが確認できたため、本研究ではこのマイクを用いた。マイクをイヤホン部分に固定するため、3D プリント (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 3c)。

なお、利用する周波数領域が 22kHz までで良い場合は、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



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

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伴い、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

提案手法を Android アプリケーション上で実装した (図3d)。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、スライダ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96kHzとした。FFTのサンプル数は1024とし、スライディング幅は512である。

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、聴力自在化のためのフレームワークの提案であるが、提案手法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。

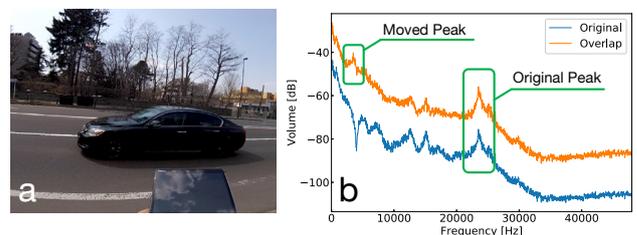


図4 (a) 音を取得したEV/HVの外観 (b) 対応する周波数スペクトル

5.1 EV/HVの検出

EVやHVは、低速時に静音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]。提案手法を用いれば、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、overlapが最適だと考えられる。overlapでは、可聴域の音は通常通りに維持しながら、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は道路の側のような、可聴音が安全のために重要な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4に示す。この図に示すように、25kHz付近に特徴的な超音波のピークが確認できる (図4b)。20kHzから40kHzを0Hzから20kHzへと重畳させることで、25kHz付近のピークが5kHz付近にも現れる。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。

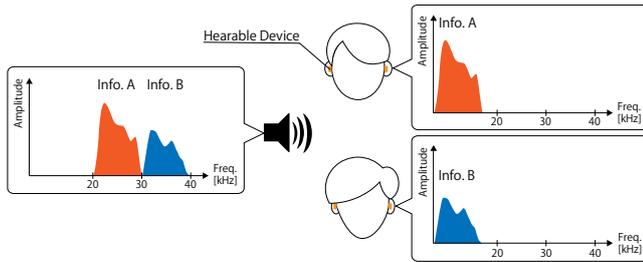


図 5 超音波音声通信

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、ガス等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聴取しやすい状態に操作できる. 以上より、提案手法を用いれば、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.

5.3 超音波音声通信

提案手法によって、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一つのスピーカから提示できる. これは、先行研究 [20] で提案されている手法と同様である. 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 5 に示すように、例えば、20k-30kHz には情報 A を、30k-40kHz 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、

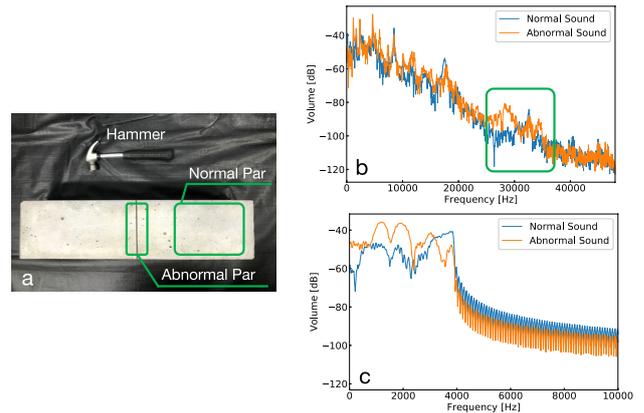


図 6 (a) 使用したコンクリートとハンマ (b) 操作前の周波数スペクトル (c) 操作後の周波数スペクトル

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 [13]. 提案手法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、実施が容易で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に一定の経験とスキルが求められる [21]. 提案手法を用いれば、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 6 に示す. 図 6a 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、100mm × 100mm × 400mm であり、1.2mm × 50mm 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 6b 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、30kHz 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 6c に示すように、slide によって 27k-31kHz を 0-4kHz 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

提案手法を用いることで、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

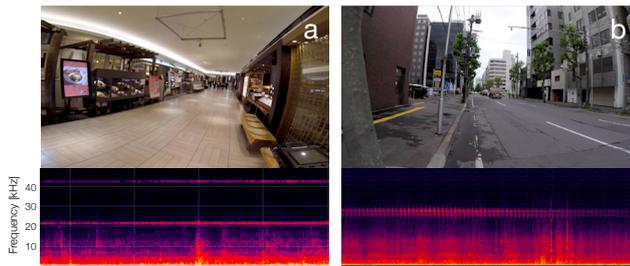


図7 (a) レストラン付近で確認できた超音波 (b) 道路付近で確認できた超音波

ンには partial elimination/extraction が利用できる。同様の利用方法は Here One[10]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るように自由にイコライズ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[12]、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。提案手法は、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、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[22]。

実際に街の中で取得できた超音波を図7に示す。図7aはレストラン付近で得られた超音波であり、20kHzと40kHz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7bは道路で取得した超音波であり、25kHz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。提案手法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. 考察

6.1 音源方向の保存

提案手法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[23]が、提案手法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8に示す。実験は3m×3mの領域で行なった。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、ランダムでどれか一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なった。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、440Hzの純音 (S1)、0-20kHzの帯域制限ノイズ (S2)、0-20kHz帯域の可聴音のアナウンス (S3)、20440Hzの純音 (S4)、20k-40kHzの帯域制限ノイズ (S5)、20k-40kHz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、指定帯域の全周波数を含む帯域制限ノイズ及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なった。ユーザが超音波 (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440Hzに、S5とS6は0-20kHzへと変換された。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及び図9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、全ての被験者は正しい音源を特定することができた。図9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。全ての音源において、音源特定に必要な時間は周波数操作後の方がかかっている。しかし、分散分析の結果、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10に、音源特定までに最も時間のかかった被験者BのS4の時の軌跡を示す。原点とx軸、y軸は図8に対応しており、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

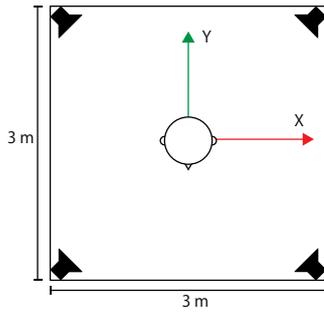


図 8 実験環境

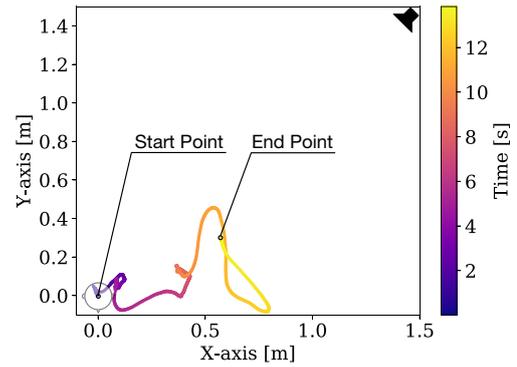


図 10 被験者 B の S4 の時の軌跡

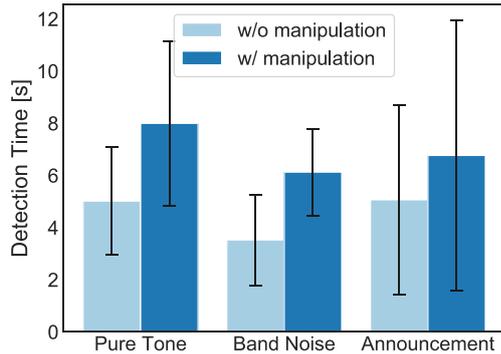


図 9 周波数操作無し/ありの結果の比較

表 1 音源特定までにかかった時間 [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

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、全ての被験者が、「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。提案手法では、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、両耳音量差 (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手法の性能で問題ないと考えられる。

表 2 各 FFT サイズにおける計算時間 [ms].

Size	Normal	Compression	Overlap	Slide	PEL ¹	PEX ²
512	5.2	5.3	5.3	5.3	5.3	5.3
1024	10.7	10.7	10.7	10.7	10.7	10.7
2048	21.3	21.3	21.3	21.3	21.3	21.3
4096	42.7	42.7	42.7	42.7	42.7	42.7

¹Partial Elimination ²Partial Extraction

6.2 周波数操作後の遅延

提案手法では、ソフトウェア上で周波数操作を行なっているため、変換による遅延が生じる。ユーザの視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。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 [24]、本研究では周波数操作を行なっている。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。提案手法では、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。そこで、本調査では 4 つの FFT サイズ (512, 1024, 2048, 4096) について調査した。

まず、各 FFT サイズにおいて、データの読み込み、周波数操作、データの出力にかかる計算時間を測定した。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。この表に示すとおり、各操作方法で計測時間に差は見られなかった。これは、どの操作方法も、FFT を計算して周波数領域で処理を行い、IFFT を行うという同様な手順を行なっているためである。また、FFT サイズと計算時間には比例に近い関係が確認できた。

次に、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 (タイピング、動画の視聴、筆記、物を叩く等) を行なっ

Score	Impairment
5	劣化が全く認められない
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

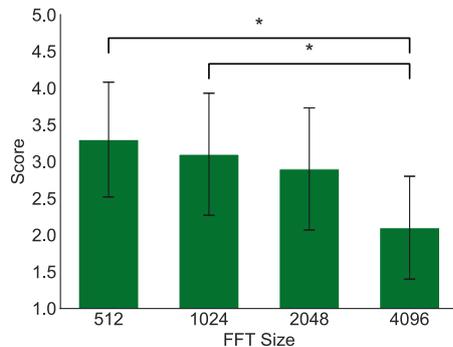


図 11 各 FFT サイズにおける遅延の結果

てもらう。被験者は、自分の行な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。評価指標は 1 から 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 3 に示す尺度で評価を行なった。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 11 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、1024 で 3.1、2048 で 2.9、4096 で 2.1 となった。分散分析の結果、FFT サイズによる有意差が確認できた ($p < 0.01$)。Bonferroni 法による多重比較を行なった結果、512 と 4096、1024 と 4096 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。しかし、FFT サイズが 2048 以下の場合、有意差は確認できなかった。

最も高いスコアでも、FFT サイズが 512 の時の 3.3 であり、これは十分に高い値とはいえない。しかし、被験者 3 からは「FFT サイズが 4096 の時、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、ユーザは遅延に適應できる可能性がある。長時間装着による影響の変化の調査は今後の課題としたい。

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 [19]。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、普段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 21kHz から 29kHz に、空港でのアナウンスを 30kHz から 38kHz 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 0kHz から 8kHz へと遷移することで、可聴音に復調する。周波数分解能は FFT サイズによって変化する。そのため、512、1024、2048、4096 の 4 通りの FFT サイズについて音質を調査した。音源は全て被験者の母国語 (日本語) で作成した。評価指標として、Degradation Mean Opinion Score (DMOS) を用いた [25]。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 3 に示す尺度に基づいて行われた。被験者は 22-29 歳の男性 9 名と女性 1 名である。

6.3.2 結果

図 12 と図 13 に結果を示す。図中のエラーバーは標準偏差を示す。21k-29kHz で変調した場合、FFT サイズが 512 では 2.2、1024 では 3.8、2048 では 3.4、4096 では 4.1 となった。30k-38kHz で変調した場合、FFT サイズが 512 では 1.8、1024 では 2.9、2048 では 2.6、4096 では 3.5 となった。分散分析の結果、20kHz 台を用いた際にも、30kHz 台を用いた際にも、FFT サイズによる有意差が確認できた ($p < 0.01$)。Bonferroni 法による多重比較の結果、20kHz 台の場合、512 と 1024 の間、512 と 2048 の間、512 と 4096 の間に、それぞれ有意差が確認できた ($p < 0.05$)。また、30kHz 台の場合、512 と 1024 の間、512 と 2048 の間、512 と 4096 の間、2048 と 4096 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悪く、FFT サイズが 4096 の時に最もスコアが良くなった。21k-29kHz と 30k-38kHz を用いた際の結果を比較すると、21k-29kHz を用いた結果の方が良いスコアが得られていた。これは、高周波の方が空気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、全被験者が全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、提案手法による超音波音声通信アプリケーションは実現可

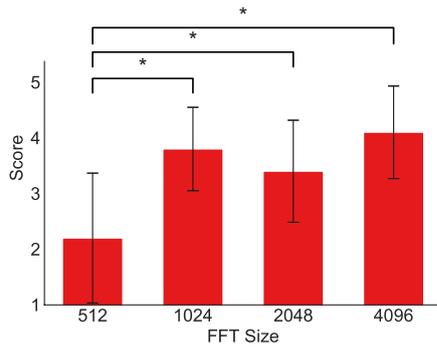


図 12 20kHz 台を用いた時の各 FFT サイズにおける音質の結果

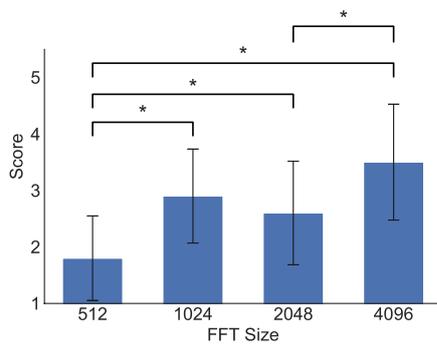


図 13 30kHz 台を用いた時の各 FFT サイズにおける音質の結果

能だと考えられる。

6.4 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 14 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 1024 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 1024 が最適だといえる。

6.5 課題と今後の発展

6.5.1 直接音の影響

遅延による不快感の原因の一つが、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

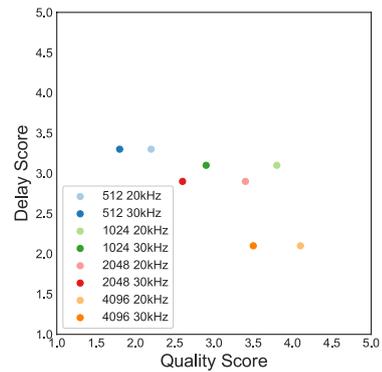


図 14 遅延と音質の結果

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。

6.5.2 デバイス性能

本研究では 96kHz のサンプリングを行なったが、一般的なヒア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ブルデバイスのみで提案手法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/ウェアブルデバイスを組み合わせることで、提案手法は実現可能である。今後の機器の発展に伴い、ヒアブルデバイスのみでの提案手法の実装が可能になると考える。

6.5.3 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、打音検査等をあげたが、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全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。提案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、今後調査を行なう必要がある。

6.5.4 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。

6.5.5 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、提案手法は時間や空間方向への応用も考えられる。例えば、取得し

た音声の再生速度の操作や、特定方向の音源の音量のみを強調するといった操作も考えられる。時間と空間に対する聴力の操作技術の調査は、今後の課題である。

7. 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、提案手法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、周波数操作に伴う音の遅延と音質のトレードオフについても議論した。提案手法は、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. of the 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. of the 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. of the 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, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pp. 267–275 (2014).
- [5] 3DSignals: Sound-based Predictive Maintenance (2019). <https://www.3dsig.com/>.
- [6] Emerson: About Ultrasonic Gas Leak Detectors (2019). <https://www.emerson.com/en-sg/automation/measurement-instrumentation/flame-gas-detection/about-ultrasonic-gas-leak-detectors>.
- [7] Pettersson Elektronik: Bat detectors and sound analysis software (2019). <http://www.batsound.com/>.
- [8] ReSound: Sound Shaper - Newest hearing aids tailored to your individual hearing environment and preference (2019). <http://www.gnhearing.dk/en/professionals/technology-and-innovation/surround-sound-by-resound/sound-shaper>.
- [9] Oticon: Inium Sense - Speech Rescue™ technology (2019). <https://www.oticon.com/professionals/brainhearing-technology/brainhearing-technology/speech-rescue>.
- [10] Doppler Labs: Here One (2019). <https://hereplus.me/products/here-one/>.
- [11] Bose: Bose Headphones (2019). https://www.bose.com/en_us/products/wellness/conversation_enhancing_headphones/headphones.html.
- [12] 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. of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 2943–2952 (2015).
- [13] 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).
- [14] 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. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pp. 1333–1338 (2015).
- [15] Ando, T., Kubo, Y., Shizuki, B. and Takahashi, S.: CanalSense: Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals, *Proc. of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 679–689 (2017).
- [16] 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).
- [17] NEC: Human-Oriented IoT Solutions Using Hearable Technology from NEC (2019). <https://www.nec.com/en/global/techrep/journal/g17/n01/170110.html>.
- [18] Amesaka, T., Watanabe, H. and Sugimoto, M.: Facial Expression Recognition Using Ear Canal Transfer Function, *Proc. of the 23rd International Symposium on Wearable Computers*, ISWC '19, pp. 1–9 (2019).
- [19] Laroche, J. and Dolson, M.: New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects, *Proc. of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, WASPAA '99, pp. 91–94 (1999).
- [20] 吉野祥之, 苗村 健: u-soul: 超音波を用いた空間的な音像提示システム, *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).
- [21] Yamashita, A., Hara, T. and Kaneko, T.: Inspection of Visible and Invisible Features of Objects with Image and Sound Signal Processing, *Proc. of the 2006 IEEE International Conference on Intelligent Robots and Systems*, pp. 3837–3842 (2006).
- [22] 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).
- [23] Blauert, J.: *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press (1996).
- [24] Alexander, J.: Hearing Aid Delay and Current Drain in Modern Digital Devices, *Canadian Audiologist*, Vol. 3, No. 4 (2016).
- [25] International Telecommunication Union: P.800 : Methods for subjective determination of transmission quality (2019). <https://www.itu.int/rec/T-REC-P.800-199608-I/en>.