



Title	機械学習を用いた超音波通信システムに関する研究
Author(s)	尾関, 剛成
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16011号
Issue Date	2024-03-25
DOI	https://doi.org/10.14943/doctoral.k16011
Doc URL	https://hdl.handle.net/2115/92316
Type	doctoral thesis
File Information	Kosei_Ozeki.pdf



博士後期課程 学位論文

機械学習を用いた超音波通信システムに
関する研究

A study on ultrasonic signals communication systems
using machine learning



尾関 剛成

北海道大学 大学院 情報科学科 情報科学専攻
メディアネットワークコース
情報メディア環境学研究室

2023年3月

目次

第1章	序論	1
1.1	研究の背景	1
1.2	研究の経緯	3
1.3	研究の目的	5
1.4	論文の構成	6
第2章	関連研究	7
2.1	音波通信	7
2.2	関連する従来手法	8
2.3	機械学習による音響処理	9
第3章	超音波通信システム	11
3.1	超音波通信	12
3.1.1	超音波通信の特性	12
3.1.2	通信方式	12
3.1.3	音波信号	14
3.2	従来の超音波通信システム	19
3.2.1	従来の超音波通信システムの設計と機能	19
3.2.2	音波信号の配置の決定	23
第4章	従来技術の実装	25
4.1	雑音特性に対する精度と課題	25
4.2	周波数特性に対する精度と課題	29
第5章	プロトタイプの開発	34
5.1	簡易音波信号	34
5.2	前処理	35
5.3	プロトタイプで設計する機械学習モデル	37
5.4	プロトタイプにおける機械学習	38
5.5	簡易音波信号の録音	39

5.6	プロットタイプにおけるファインチューニング	41
5.7	プロットタイプの評価	43
5.8	プロットタイプの開発の考察・まとめ	45
第 6 章	提案する超音波通信システム	46
6.1	超音波通信システムの設計	46
6.2	実験の設計	50
6.2.1	ホッピングパターン予測器の設計	51
6.2.2	周波数シフト予測器の設計	54
6.3	データセットの収集	54
6.4	提案予測器の学習および性能	57
6.4.1	ホッピングパターン予測器の学習評価	57
6.4.2	ホッピングパターン予測器の性能評価	62
6.4.3	周波数シフト予測器の学習評価	65
6.4.4	周波数シフト予測器の性能評価	72
6.5	性能評価	77
第 7 章	まとめと今後の展望	82

概要

近年、スマートフォンやタブレットなどのモバイル端末の普及が進み、さまざまな場所で簡単に情報にアクセスできる時代になってきた。しかし、デファクトスタンダードとして利用されている電波通信は、技術の進歩に伴うアップデートが継続的に行われており、非対応の端末が混在する問題が避けられないことや、水中や地下、金属構造の建物内などの非電波透過環境や病院・工場などの電波通信の使用が制限される環境では通信が困難になることから、社会インフラとしての運用には発展途上の課題が残されている。

他方、こうした隙間をカバーする通信方式として期待されているのが音波通信である。音波通信は、空気の振動により情報を送受信するため、水中などの電波が届かない範囲でも通信を行うことができる。超カバレッジ拡張を目的とした第6世代移動通信システムの展開を想定すると、音波通信技術の発展は情報社会にとって、非常に大きい影響力を有することが考えられる。また、音波通信は、マイクやスピーカなどの音声入出力機能を利用するだけで実現できるため、店内に設置されているスピーカや家庭用テレビから発信される音響に情報を埋め込むだけで通信が利用でき、導入コストが低いことも特徴の一つである。

一方、音波通信は、電波通信と比べると環境雑音や反響音、残響音、ドップラー効果などの外乱により音波信号が歪みやすく、情報識別・信号同期に誤りが生じやすいことが問題になっている。これらの課題に対する改善策として、雑音や周波数シフトに応じて処理を調整する方法や音量を増加させて信号を送信する方法が考えられるが、環境に適応した調整は手間であることや人体への影響を懸念すると現実的ではない。

こうした問題に対する解決方策として、本研究では、機械学習に基づく最新の情報識別技術に着目した。機械学習により設計された識別器は、外乱によって劣化した情報に対しても高い識別性能を示すことが報告されており、画像認識問題や文字認識問題、音響分離問題等広い分野にて適用されている。

本研究では、非可聴域である高周波帯域に情報を埋め込み、白色雑音と周波数シフトにより劣化した音波信号をシミュレーションし、CNN、RNN、CRNNによる機械学習モデルを用いて外乱下から音波信号の同期処理および音波信号にかかる周波数シフトの値を予測する識別器を設計し、その有効性を示す。

機械学習では、音波信号を時間と周波数の情報を2次元の画像として表現するスペクトログラムを学習データとして用い、スペクトログラムのパターン情報から音波信号の特徴を抽出するように設計し、雑音環境下および周波数シフト発生時における音波信号のBERの識別精度の性能評価を行った。

結果、従来のバンドパスフィルタ処理を用いる超音波通信システムでは、BER=0%を示す音波信号であっても、25Hz以上の周波数シフトが発生するとBERは20.82%を上回り、信号識別の精度が大きく低下する。一方、CRNNを用いた機械学習モデルは、入力されたスペクトログラムから音波信号に含まれる周波数シフトの値を予測することにより、SNR=6dB以上の環境下において、50Hzの周波数シフトが発生してもBER=0%を示した。

本研究で提案する機械学習を用いた識別器は、従来の超音波通信システムで対応が困難であった周波数シフトの影響を最小限に抑え、高い識別精度が得られることを明らかにし、その有効性を示した。

また、基礎研究にとどまらず、実際にサービスを展開している企業と連携し、産業応用という出口を確保しながら研究を推進していることが本研究の特色のひとつになっている。

第1章

序論

ここでは、序論として、本研究の背景およびその経緯と目的、本論文の構成について述べる。

1.1 研究の背景

近年、スマートフォンやタブレットなどのモバイル端末の普及が進み、さまざまな場所で簡単に情報にアクセスできる時代になってきた。しかし、デファクトスタンダードとして利用されている電波通信は、技術の進歩に伴うアップデートが継続的に行われており、非対応の端末が混在する問題が避けられないことや、アンテナ装置などの専用機器が必要であること、水中や地下、金属構造の建物内などの非電波透過環境や病院・工場などの電波通信の使用が制限される環境では通信が困難になることから、社会インフラとしての運用には発展途上の課題が残されている。

他方、こうした隙間をカバーする通信方式として期待されているのが音波通信である。音波通信は、音声入出力機能を利用するだけで実現できるため、電波通信による情報のやり取りが困難な状況であっても運用が可能という利点を有する。また、2030年代の実現をめざす第6世代移動通信システム(以下、6G)に向けて注目されている技術でもある。6Gとは「Beyond 5G」とも呼ばれ、現在普及が進んでいる5Gの性能をさらに進化させた次世代の移動通信システムのことを指し、6Gの実用化は超高速、大容量、高信頼、低遅延、通信エリアの拡張、同時多数接続といった通信の高度化を実現でき、さまざまな社会変革が起きると予想されている。音波通信はその中でも、陸に加えて海・空・宇宙など、あらゆる場所での通信エリアの拡張を目指す超カバレッジ拡張において注目されている技術である。音波は、水中や地下、金属構造の建物内などの電波が届かない環境でも伝搬するため、海中無線通信においては、長距離な通信が可能な海中音響通信が注目されている。

電波通信と音波通信を利用できる環境の例を表 1.1 に示す。

表 1.1 電波通信と音波通信の利用環境

環境		通信媒体	電波通信	音波通信
	室内		○	○
	屋外		○	○
	水中		△	○
	雨天・雪天		○	△
特殊な環境	鉱山・地下空間		△	○
	電波が使えない病院内部		×	○
	電波が使えない工場内部		×	○

海中音響通信の他にも、通常の聴覚範囲を超える高い周波数の超音波を用いた超音波通信などの注目度も上がっている。超音波通信は、通信が目立たず周囲の環境に影響なく利用が可能であることや、数センチから数十メートルの距離で高い精度での情報伝送が可能である。実際に図 1.1 に示すようなスマートフォンのマイクやスピーカを用いた決済アプリケーションの開発[1]や、テレビや店内スピーカなどの既存のデバイスからクーポン情報を発行するサービスの開発[2]により社会インフラとしての運用にも用いられている。

また音波通信は、マイクやスピーカなどの音声入出力機能を利用するだけで実現できるため、店内に設置されているスピーカや家庭用テレビから発信される音声に情報を埋め込むだけで通信が利用でき、導入コストが低いことも特徴の一つである。その例として、オムロンヘルスケア株式会社では、コロナ禍において従来の体温計のハードの設計を変えずに、音波により測定した体温のデータをスマホなどに転送する無線通信を短期開発している。[3]

本研究では、株式会社スマート・ソリューション・テクノロジー(以下 SST)[1]との共同研究として行われ、実際のサービスへの応用や組み込みを考慮した超音波通信システムを設計する。

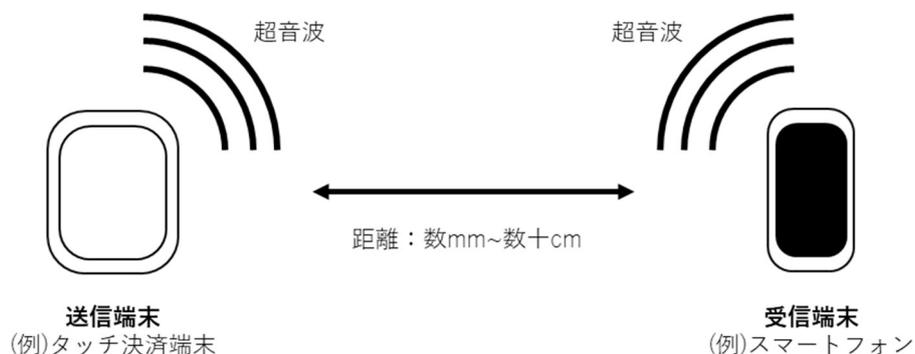


図 1.1 超音波通信システムの一例

1.2 研究の経緯

前節で述べたように、音波通信にはさまざまな利点があるものの、電波通信と比べると、使用可能な周波数帯域が狭く、通信速度が遅いため一般的な通信手法としては用いられない。また、環境雑音や反響音、残響音などの外乱により音波信号が歪みやすく、情報識別・信号同期に誤りが生じやすいことが問題になっている。電波通信と音波通信の特性の違いを表 1.2 に示す。

表 1.2 電波通信と音波通信の特性の違い

特性 \ 通信媒体	電波通信	音波通信
伝搬距離	数cm ~ 数百km以上	数cm ~ 数十m
周波数帯域	数kHz ~ 数GHz	数Hz ~ 数十kHz
通信速度	100kbps ~ 数Gbps以上	数bps ~ 数Mbps
導入コスト	比較的高い	低い
通信障害の例	遮蔽物、 電磁干渉、多重反射、 信号衰弱、ジャミング	遮蔽物、 音の吸収・反射・散乱、 騒音や環境音の影響、 周波数シフト・ドップラー効果

図 1.1 に示す決算アプリケーションなどの近距離通信は、音源との距離が近いため送信信号をクリアに録音しやすく誤りが発生しにくいですが、図 1.2 に示す店内スピーカなどを用いた数 m から数十 m の通信では、録音時に環境雑音が入り込みやすいことや、通信媒体である超音波が空気中での伝播時に減衰しやすいことから、通信の品質が低下しやすい。

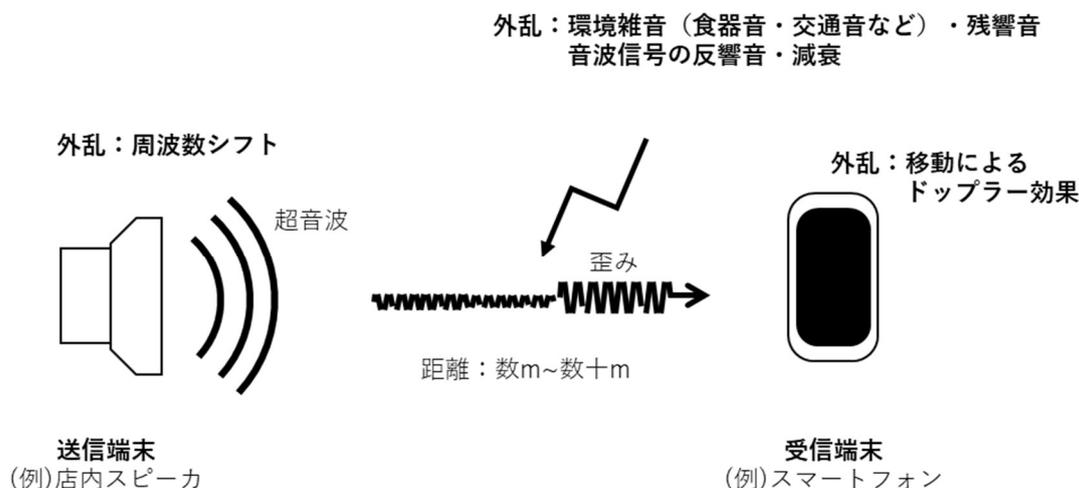


図 1.2 超音波通信システムとその障害となる外乱

環境雑音や残響音は、送信端末を設置する環境において発生する雑音であり、食器や会話などの音が特定の周波数帯域の送信信号に加算されることで、受信信号の検出を困難にする特性を持つ。

また、音波通信は電波通信と比べると、使用できる周波数帯域が狭いため、周波数に影響を与えるドップラー効果や周波数シフトなどの外乱影響を受けやすい。ドップラー効果は、受信端末が送信端末に対して移動する場合に生じる現象であり、受信端末が送信端末に接近すると、波の圧縮が生じ受信信号の周波数が高くなり、遠ざかる場合には波の伸張が生じ、受信信号の周波数が低くなる。周波数シフトは、通信システムが外部からの電磁的な干渉やノイズにより、信号に変調が生じ受信信号の周波数が変化する現象である。特に音波通信は、マイクとスピーカだけで通信ができるため、機器やインフラの設置が容易なため、低コストな通信システムが構築できるが、この簡易さと低コスト化の反面、外部からの影響に弱く周波数シフトが生じやすいという特徴もある。このドップラー効果および周波数シフトが原因により受信信号の周波数が変化し、通信の品質が低下することがある。

これらの課題に対する改善策として、商品ごとに雑音や周波数シフトに応じてフィルタ処理による雑音抑制や通信に用いるキャリア周波数を微調整する方法や、送信信号の音量を増加させる方法が考えられるが、環境に適応した調整は手間であることや人体への影響を懸念すると現実的ではない。

一方、近年は深層学習による機械学習理論が発展し、GPU やクラウド計算基盤などの計算機環境が普及したことで、音声や楽音に限らないあらゆる音を直接分析する研究が増加しており、環境音の解析や雑音抑制、音源分離など、新たな音響情報抽出の手法が提案されている。そこで、環境雑音や反響音、残響音、ドップラー効果や周波数シフトなどの外乱が及ぼす音波信号に対する問題の解決策として、本研究では、機械学習に基づく最新の情報識別技術に着目した。機械学習により設計された識別器は、外乱によって劣化した情報に対しても高い識別性能を示すことが報告されており、画像認識問題や文字認識問題、音響分離問題等広い分野にて適用されている。

1.3 研究の目的

本研究では、非可聴域である 18kHz~20kHz の高周波帯域に情報を埋め込み、通信を行う超音波通信システムにおいて情報識別・信号同期に誤りを生じさせる原因である環境雑音、ドップラー効果や周波数シフトなどの外乱に対して有効的な機械学習アプローチを提案する。

提案する機械学習アプローチとして、白色雑音とドップラー効果、周波数シフトにより劣化した音波信号をシミュレーションし、Convolutional Neural Network (以下, CNN), Recurrent Neural Network (以下, RNN), Convolutional Recurrent Neural Network (以下, CRNN) による機械学習モデルを用いて外乱下から音波信号の持つ情報を取得する識別器を設計する。

なお、本研究は、SST[1]との共同研究として行われ、提案する機械学習アプローチは、実際のサービスへの応用や組み込みを考慮して設計することを目的としている。

1.4 論文の構成

本稿は、全7章で構成される。第1章は序論である。

第2章では、関連研究として、音波通信における関連研究および環境音に対して機械学習を用いた手法について紹介する。

第3章では、本研究で扱う超音波通信システム、音波信号の定義について述べる。

第4章では、従来技術の設計方法と実装による評価実験、課題について述べる。

第5章では、機械学習モデルと音波信号の識別の有効性について、簡易的なプロットタイプの開発を行い、その設計法について述べる。

第6章では、プロットタイプにより得られた効果を元に、本研究で提案する機械学習を超音波通信システムについて述べる。本研究では、同期処理と周波数シフトの予測の2つタスクを行う予測機を設計する。その設計方法について、解説し学習データの内容や学習過程の評価およびテスト用データセットに対する結果および評価について述べる。

第7章は、本研究を総括し、今後の課題を述べる。

第2章

関連研究

ここでは、関連研究として従来の音波通信および音声・音響に対する機械学習アプローチについて説明する。

2.1 音波通信

音響信号を介してデジタル情報を伝送する技術は、一般的に「空間伝搬音響情報伝送 (aerial acoustic data transmission) 技術」として知られている。しかし、以降、本論文では電波通信と対比して「音波通信(sonic data transmission)」と簡略化して呼ぶこととする。以下、音波通信の特性や用途、また、その利点・欠点について概説する。

音波通信は、ごく近距離から数百メートルの範囲で利用される。特に超音波通信は、数センチから数十メートルの距離で高い精度での情報伝送が可能である。この特性から、音波通信は室内測位、近距離決済、およびインタラクティブなユーザーエクスペリエンスの向上など、様々な応用分野で利用が期待されており、音波を活用した近・中距離通信技術の可能性が検討されてきた[1,2,3,4]。この音波通信は、通話機能を備えたモバイル端末に搭載されている音声の入出力機能（マイクとスピーカ）を活用することで、容易に実現できる利点がある。最近では、この技術を用いて伝送された情報を、ユーザーの手元の携帯情報端末で復号・表示させ、様々な体験を実現するための応用が活発に行われている。しかし、音波通信は、電波通信と比べると、使用できる周波数帯域が狭いため、周波数に影響を与えるドップラー効果や周波数シフトなどの影響を受けやすいことや、環境の雑音や反響音、残響音などの干渉により信号波形が歪みやすく、信号検出において外乱の影響が大きいなどの問題も抱えていることが認識されている[5,6,7]。

音波通信技術は、2030年代の実現をめざす6Gの実用化に向け陸に加えて海・空・宇宙など、あらゆる場所での通信エリアの拡張を目指す超カバレッジ拡張において注目されている技術である。音波は、水中や地下、金属構造の建物内などの電波が届かない環境でも伝搬するため、海中無線通信においては、長距離な通信が可能な海中音響通信が注目されており、情報社会にとって、非常に大きい影響力を有することが考えられる[8]。超カバレッジ拡張の実現により電波が届かない範囲でも通信を行う環境が整うことで、多様な産業を支える社会インフラとしての役割が期待され、新たな応用分野での展開が可能となりの様々な側面において先進的な通信技術がもたらす革新的な変化に貢献することが期待されている。

本研究では、通話機能を備えたモバイル端末に搭載されている音声の入出力機能（マイクとスピーカ）を活用することで、比較的実装が容易な音波通信を扱う。音波信号は、非可聴域である18kHz~20kHzの高周波帯域の超音波を扱い、情報識別・信号同期に誤り生じさせる原因である環境雑音や反響音、残響音、ドップラー効果や周波数シフトなどの外乱に対して有効的な機械学習アプローチを提案する。

2.2 関連する従来手法

先行研究では、音メディアを介したデータ通信を実現するためのさまざまなシステムが提案されている[4,5,6,7]。これらの中には、高周波帯域にバイナリ情報を埋め込む超音波通信システムを採用しているものもある[1]。この手法により、通信が視覚的に目立たず、かつ人間の耳にはほとんど感知されないでデータを伝送することが可能となる。その結果、情報の機密性が向上し、通信のセキュリティが強化され、特に機密情報やプライバシーに関連するデータの送信がより安全かつ効果的に行えると期待される。このような音波通信システムは、Band-pass filter（以下BPF）、Low-pass filter（以下LPF）、により特定の周波帯域の音波信号を抽出し、送信信号と受信信号の相関係数を用いることで同期処理および解析、バイナリ情報の照合を可能とする。従来の超音波通信システムの具体的な処理方法については次章で述べる。

2.3 機械学習による音響処理

本研究では、最新の機械学習に基づく情報識別技術を音波通信に応用する。機械学習によって設計された識別器は、雑音などの外乱による情報の劣化に対して高い識別性能を示すことが報告され、画像認識問題[9]や文字認識問題[10]、音響分離問題[11]など、広範な分野での応用が行われている。

2010年代以降、深層学習による機械学習理論が発展し、GPUやクラウド計算基盤などの計算機環境が普及したことで、音声や楽音に限らないあらゆる音を直接分析する研究が増加した。これにより、環境音、会話、音楽、異常音にまで注目が向けられ、新たな音響情報抽出の手法やアプリケーションが次々と提案されるようになった。機械学習を音声処理に用いた代表的な例として、発話音声をテキストに変換する音声認識[12,13]、与えられた音声から話者を識別する話者識別[14,15]、テキストから自然な音声を生成する音声生成などが挙げられる[16,17]。これら以外にも、さまざまな音声処理への機械学習の導入が進んでいる。例えば、機械学習を環境音に適用する際の役割として、入力音を分析し、特定の音響パターンや状況を認識し、音響データの理解や処理を行う音響シーン分類[18]や音響イベント検出[19,20]、異常音検知[21,22]などのタスクが挙げられる。

音響シーン分類の目的は、入力音から収録された場所や状況を特定することにある。具体的な例としては、電車、車、公園、屋内などの異なる音響シーンラベルを出力することが挙げられる。機械学習モデルは、入力音の特徴を学習し、これらの異なる音響シーンに関連付けられたラベルを出力することで、環境の種類や背景状況の理解が可能となる。音響イベント検出の目的は、入力音から足音、車の走行音、ドアの開け閉めの音など、細かい音のラベルと発生区間を特定することにある。機械学習モデルは、異なる音響イベントのパターンを学習し、これらが入力音に現れる場合に対応するラベルと発生区間を出力する。これにより、具体的な音響イベントの検出が可能となる。異常音検知の目的は、通常的环境音とは異なるパターンを持つ音を検出することにある。機械学習モデルは環境音のパターンを学習し、学習したパターンとは異なる音を異常として検出する。異常音検知は、セキュリティや故障検知などの用途において重要であり、環境の変化や異常を素早く検知することが求められている。

また、近年は自然言語処理の分野において、特に注意機構（Attention Mechanism）の導入が大きな注目を集め、その効果が顕著に示されている。注意機構は、入力データの中で重要な部分に注意を向け、その情報に対して重み付けを行う。これにより、モデルは問題に対する特定の要素に注目し、他の部分よりも優先的に処理することで畳み込みや再帰構造よりも柔軟かつ長期の依存関係の学習がしやすいことが特徴であり、タスクの効率的な解決や精度向上に寄与するなどの効果が得られている[23]。注意機構を用いることで生成する際の精度や自然さが向上し、音声・音響処理におけるタスクでも革新的な進展がみられている[21,24]。

しかし、機械学習モデルは通常、大規模なパラメータを持ち、学習や推論の処理に高い計算リソースを要求する。これは高い精度を達成する一方で、実際のシステムに機械学習モデルを組み込む場合、処理速度の観点からは課題が生じる。そのため、機械学習モデルを実用するシステムに組み込む場合は処理速度向上のために簡易的なパラメータの少ない軽量なモデルを選択することが重要とされている。本研究では実際に超音波通信システムに組み込み、実運用を想定しているため、1～5層までの軽量なモデルを選定する。

機械学習は、雑音などの環境音を含む音波信号からデータを高度に処理し、超音波通信を効率的に利用するための手法やアルゴリズムを開発する上で、重要な役割を果たしている。超音波通信は、高い周波数の音波を使用して情報を伝送するため、信号のノイズ干渉に敏感である。これらの問題に対処し、機械学習アルゴリズムを適用することで、超音波通信は正確で信頼性の高いデータ転送を実現することができ、より高度な情報処理と柔軟性を獲得し、様々な状況において優れたパフォーマンスを発揮することが期待されている。

第3章

超音波通信システム

本章では、本研究で扱う超音波通信システムおよび音波信号の設計、従来技術の紹介およびその課題を説明する。なお、この超音波通信システムは、SST[1]にて実際に用いられる処理内容を参考に設計している。SST は超音波通信システムを用いた最新の Touch Point テクノロジーを独自に開発している会社であり、その商品例として Zeetle CS という超音波通信サービスを展開している。図 3.1 に Zeetle CS のイメージ図を示す。



図 3.1 超音波通信の実用例 (Zeetle CS)

3.1 超音波通信

3.1.1 超音波通信の特性

超音波通信は、通常の聴覚範囲を超える高い周波数の音波を使用して情報を伝送する通信方式である。「モスキートーン」で知られるように、20歳未満の若い人々は16kHzから17kHzの周波数の音を感じやすい傾向にあり、年齢が上がるにつれて聴覚が低下するため、年齢層が高い人々には聞こえにくい、ほとんど聞こえないとされている。一般的な聴覚範囲は18kHz以下であり、超音波通信はそれより高い周波数帯域を利用して、人間の聴覚範囲外で情報を伝送する通信手法である。これにより、通信が目立たず、周囲の環境に影響なく利用が可能である。また超音波は空気中での伝播時に減衰しやすいため、数センチから数十メートルの相対的に短い距離で高精度な情報伝送が可能である。

一方、超音波通信に限らず音波を用いた通信手法は、電波通信と比べて環境雑音や反響音、残響音、ドップラー効果などの外乱により音波信号が歪みやすく、情報識別・信号同期に誤りが生じやすいことや、伝送速度が低く高速なデータ伝送が制約されることがあるなど、課題も多い。

3.1.2 通信方式

本研究で扱う超音波通信システムの通信方式は周波数偏移変調（Frequency Shift Keying 以下、FSK）である。FSKは、デジタル通信における変調方式の一つで、情報を搬送波の周波数の変化によって表現する通信方式である。

最も単純な FSK はバイナリ周波数偏移変調 (BFSK) である。BFSK は、離散周波数のペアを使用して、バイナリ (0 と 1) 情報を送信する。この方式では、「0」はスペース周波数と呼ばれ、「1」はマーク周波数と呼ばれる。具体的には、図 3.2 に示すように異なるデータビットに対して異なる周波数を割り当て、これによってデジタルデータを伝送する。実際に通信を行う際は、キャリア周波数 (Carrier Frequency) にバイナリ情報を加算し、伝送される。キャリア周波数は、通信信号が伝送される際に、その信号を波として伝える基本となる周波数帯域を指す。

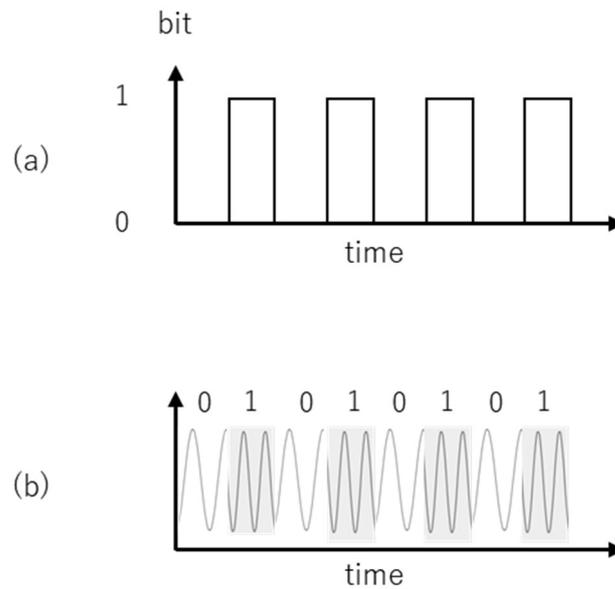


図 3.2 BFSK の例

3.1.3 音波信号

本研究で扱う音波信号の定義および概要を説明する。音波信号は次のように定義される。

$$s(t) = \exp(j2\pi(f_c + f_m)t) \quad (3.1.1)$$

f_c と f_m はキャリア周波数と変調周波数を示す。 f_c は18kHz~19.9kHzの範囲に100Hz間隔で20ch用意し、変調周波数 f_m は50Hzまたは100Hzであり、50Hzをスペース周波数、100Hzをマーク周波数として扱う。このシステムの音波信号 $s(t)$ をキャリア周波数 f_c で復調処理すると新しい信号 $s_m(t)$ は次のようになる。復調処理による変換はキャリア周波数 f_c の周波数成分を $s(t)$ から取り除くために行われ、 $s(t)$ と f_c の周波数成分との間で乗算操作を行うことで成り立つ。 $s_m(t)$ はキャリア周波数 f_c がキャンセルされ変調周波数 f_m だけをもつベースバンド信号 $s_m(t)$ として扱う。

$$s_m(t) = s(t) \cdot \exp(-j2\pi f_c t) = \exp(j2\pi f_m t) \quad (3.1.2)$$

このベースバンド信号に加算される雑音が白色雑音であると仮定すると復調した信号は次のように変換される。

$$s_m(t) = \exp(j2\pi f_m t) + n(t) \quad (3.1.3)$$

$n(t)$ は白色雑音であり、正規分布によるガウス雑音と過程する。この白色雑音 $n(t)$ とベースバンド信号 $s_m(t)$ のSN比(signal-to-noise ratio)は次のように定義される。RMSは信号の二乗平均平方根を意味する。

$$\text{SNR} = 20 \log_{10} \frac{1}{\text{RMS}(n(t))} \quad (3.1.4)$$

式(3.1.1)に示される音波信号 $s(t)$ のイメージ図を図 3.3 に示す。図中では時間軸に対する振幅の変化が視覚的に表現されており、雑音を含まない十音の音波信号によって構成されている。この音波信号の一音をシンボルと呼び、それぞれのシンボルがバイナリ情報を示している。本研究で扱う音波信号のシンボル長は一音 41.66ms とし、時間領域においてシンボル間の干渉はしないようビットレートは 24bps となるように設計している。音波信号のサンプリング周波数はナイキスト定理に基づき信号の最高周波数成分の 2 倍以上必要である。本研究では最大 20kHz の音波信号を扱うため 40kHz 以上のサンプリング周波数を設計する必要がある。一般的なスマートフォンの録音アプリでは、44.1kHz もしくは 48kHz が使用されるため、本研究の音波信号は高品質な録音が可能で 48kHz をサンプリング周波数とした。また、音波信号の変調周波数を 50Hz 間隔で設定している理由として、本研究で扱う超音波通信システムは、受信端末にスマートフォンの利用を想定しており、そのスマートフォンを $\pm 0.6\text{m/s}$ 内の速さで左右に振った場合の、 $-30\text{Hz} \sim +30\text{Hz}$ 以内のドップラー効果に対応できるように設計している。なお、 $\pm 0.6\text{m/s}$ の速さはライブ会場などでスマートフォンを振るような状況での音波通信を想定している。

図 3.3 の各シンボルに含まれるキャリア周波数 f_c は①19200Hz, ②19500Hz, ③18000Hz, ④18300Hz, ⑤19900Hz, ⑥18700Hz, ⑦18900Hz, ⑧18100Hz, ⑨19800Hz, ⑩18400Hz である。また、図 3.3 の音波信号から復調処理により得られるベースバンド信号 $s_m(t)$ のイメージ図を図 3.4 に示す。図 3.4 のベースバンド信号 $s_m(t)$ は、変調処理により音波信号 $s(t)$ のキャリア周波数 f_c がキャンセルされ、変調周波数 f_m の周波数成分のみを示す波形となっている。図 3.4 の各シンボルに含まれる変調周波数 f_m は①50Hz, ②100Hz, ③100Hz, ④50Hz, ⑤100Hz, ⑥100Hz, ⑦100Hz, ⑧100Hz, ⑨100Hz, ⑩100Hz である。

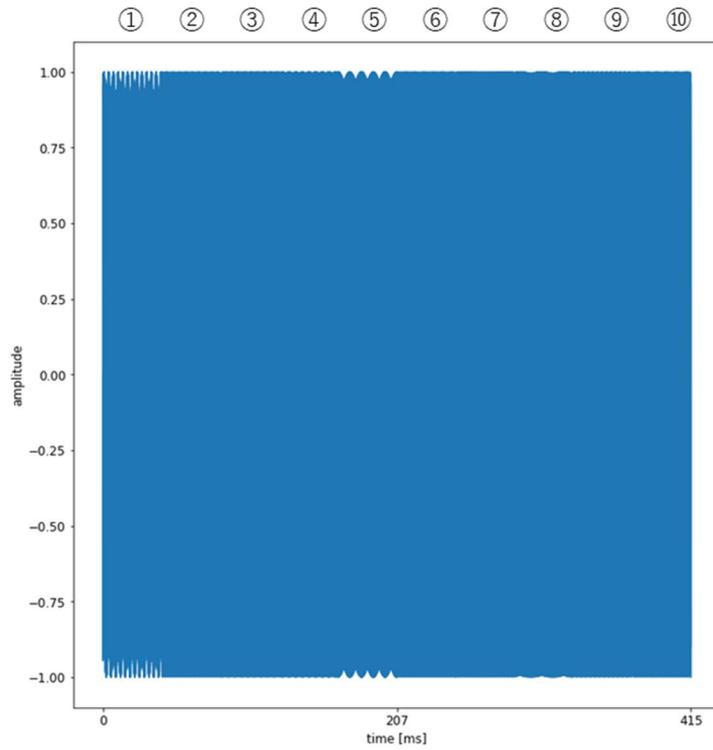


図 3.3 10 シンボルの音波信号

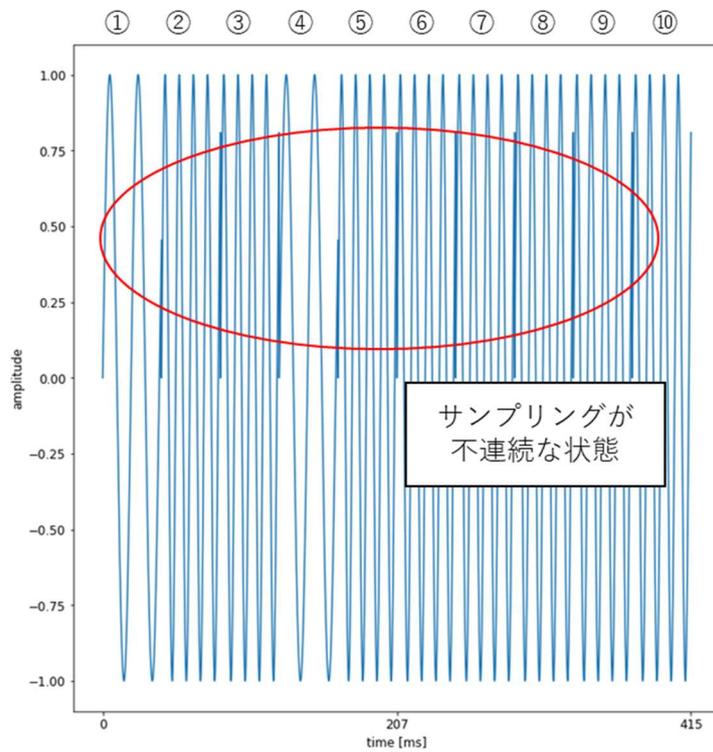


図 3.4 10 シンボルのベースバンド信号

図 3.3 の音波信号を送信するにあたり，有限の時間領域内でシンボルを切り取る操作によりシンボル間でサンプリングが不連続となるため，1 シンボルごとに図 3.5 に示すような窓関数をかける必要がある．代表的な窓関数として図 3.5 に示すハニング窓（Hanning Window）がある．ハニング窓はサイドローブの低減と，メインローブの広がり小さいことが特徴であり，音声処理やスペクトル解析で使用される．

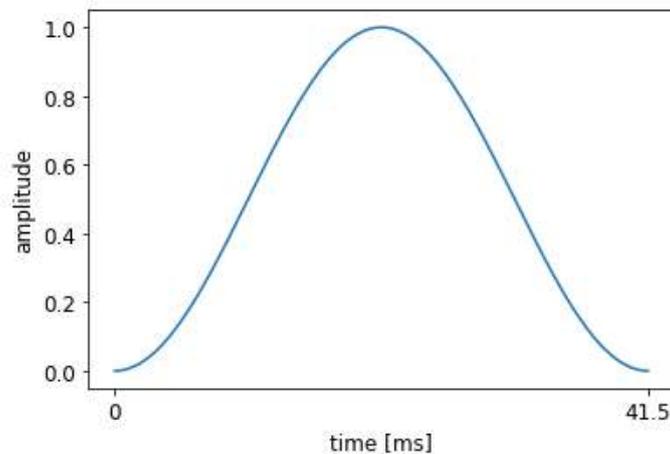


図 3.5 窓関数（ハニング窓）

窓関数とは，有限時間領域内で観測される信号において不連続性を減少させ，周波数解析や信号処理において発生する問題を軽減するための関数である．音波信号 $s(t)$ に窓関数 $w(t)$ をかけると処理後の音波信号 $x(t)$ は次のように変換される．

$$x(t) = s(t) \cdot w(t) \quad (3.1.5)$$

図 3.3 の音波信号 $s(t)$ に図 3.5 の窓関数 $w(t)$ をかけた音波信号 $x(t)$ のイメージ図を図 3.6，図 3.3 のベースバンド信号 $s_m(t)$ に窓関数 $w(t)$ をかけたベースバンド信号 $x_m(t)$ のイメージ図を図 3.7 に示す．窓関数をかけることにより，各シンボルの端は振幅値が 0 となり，シンボル間のサンプリングは連続する．本研究では音波信号に窓関数をかけた状態の波形を通信に用いる．

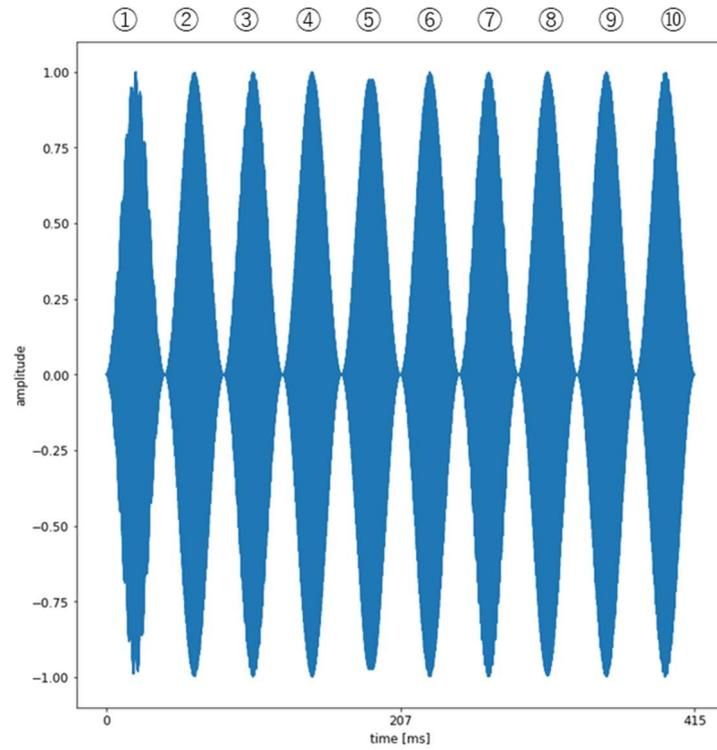


図 3.6 窓関数をかけた 10 シンボルの音波信号

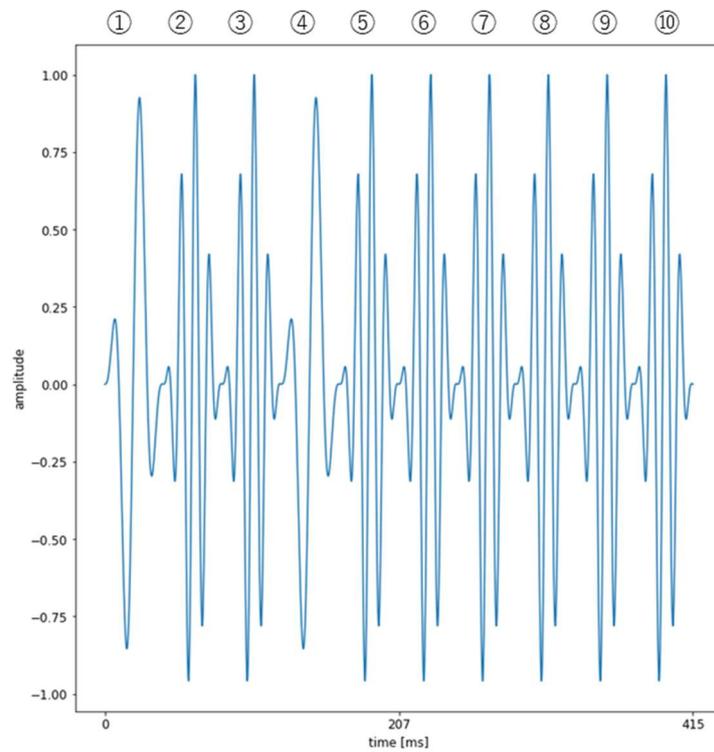


図 3.7 窓関数をかけた 10 シンボルのベースバンド信号

3.2 従来の超音波通信システム

3.2.1 従来の超音波通信システムの設計と機能

従来技術として紹介する超音波通信システムの設計について述べる。なお、この超音波通信システムは、SST[1]にて用いられている処理内容を参考に設計している。音波信号にかける窓関数は、図 3.8 に示すようなハニング窓を改良した独自の窓関数が用いられている。この窓関数は、ハニング窓に比べてメインローブの広がりが大きく、窓関数により音波信号が減衰しないように設計されていることが特徴である。この窓関数により図 3.3 の音波信号は図 3.9 に示すような波形へと変換される。この超音波通信システムでは、初めに入力された音波信号を前節で示す復調処理を行い、キャリア周波数をキャンセルすることによって、ベースバンド信号に変換する。音波信号のキャリア周波数は 18kHz~19.9kHz の範囲に 100Hz 間隔で 20ch 用意しているため、復調処理により 20ch 分のベースバンド信号に変換される。図 3.9 の音波信号を入力としたとき、超音波通信システムにより復調処理された 20ch 分のベースバンド信号を図 3.10 に示す。

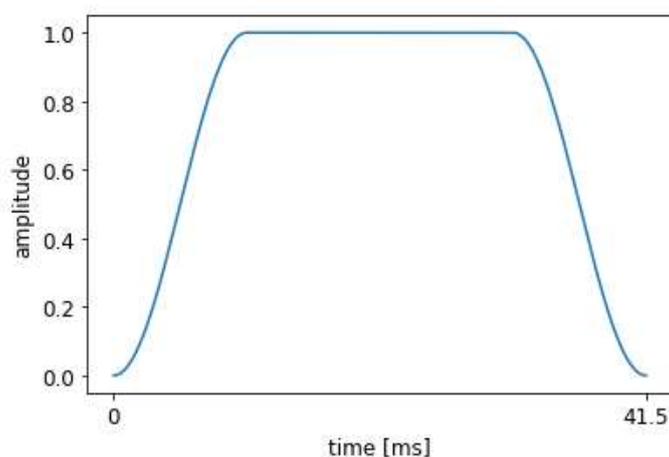


図 3.8 SST で用いられている独自の窓関数

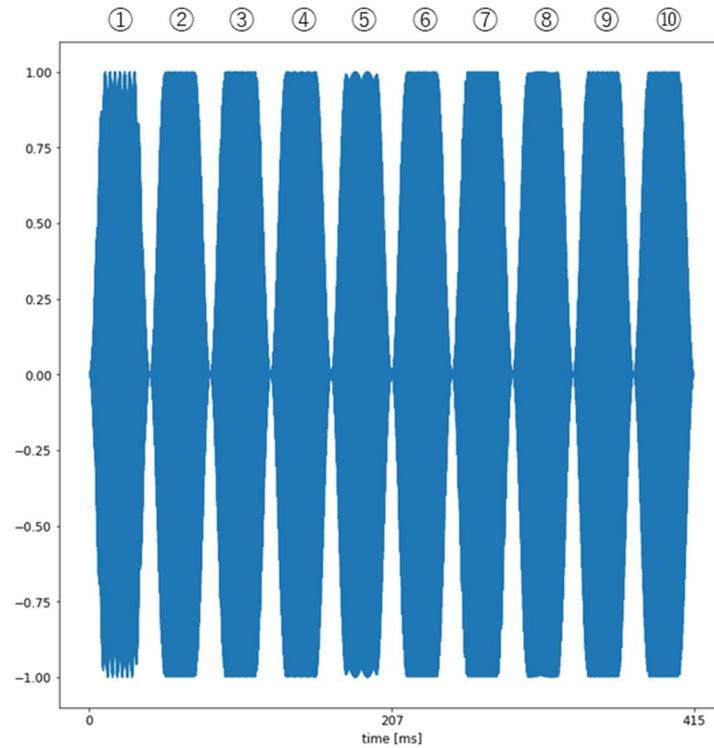


図 3.9 独自の窓関数をかけた 10 シンボルの音波信号

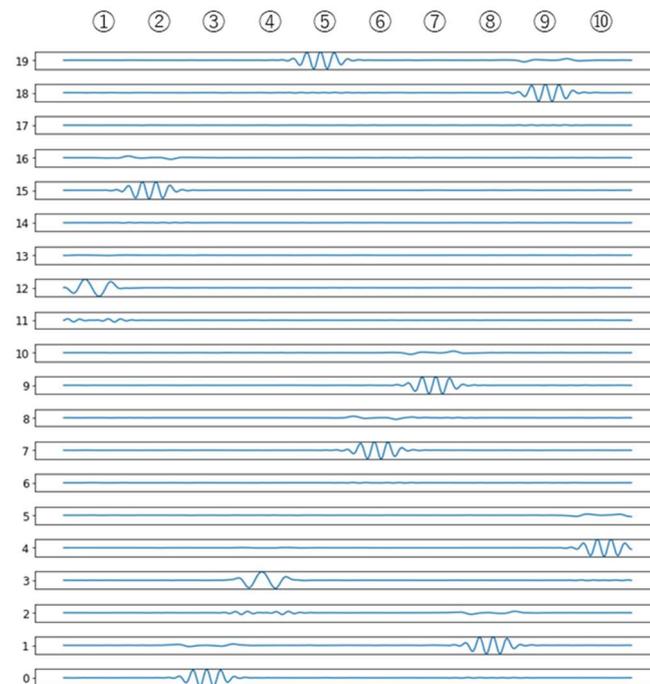


図 3.10 復調処理により 20ch に分けられたベースバンド信号

図 3.10 では 0ch (18kHz) から 19ch(19.9kHz)のキャリア周波数から得られるベースバンド信号を示している。図 3.10 の各シンボルに含まれるキャリア周波数は①19200Hz, ②19500Hz, ③18000Hz, ④18300Hz, ⑤19900Hz, ⑥18700Hz, ⑦18900Hz, ⑧18100Hz, ⑨19800Hz, ⑩18400Hzである。この復調処理後, 得られたベースバンド信号は, 50Hz のスペース周波数, および 100Hz のマーク周波数を抽出するように設計された BPF 処理により, 変調周波数が検出される。図 3.10 のベースバンド信号に対して BPF 処理を行い, 抽出された各シンボルのもつ変調周波数を示す波形を図 3.11 に示す。

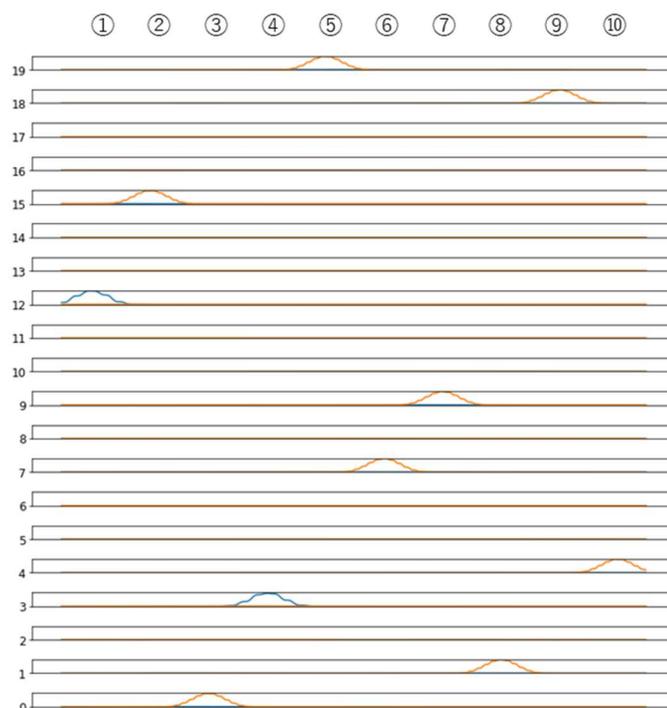


図 3.11 BPF 処理による抽出される各シンボルの変調周波数を示す波形

図 3.11 の各シンボルに含まれる変調周波数は①50Hz, ②100Hz, ③100Hz, ④50Hz, ⑤100Hz, ⑥100Hz, ⑦100Hz, ⑧100Hz, ⑨100Hz, ⑩100Hzであり, スペース周波数の 50Hz は青線, マーク周波数の 100Hz は赤線で示されている。このように従来の超音波通信システムでは, 復調およびフィルタ処理を用いることで各シンボルの変調周波数を識別することで, 音波通信を実現している。

本研究の従来技術として扱う超音波通信システムの処理概要図を図 3.12 に示す。録音された音波信号（図 3.9）は復調処理により得られた 20ch ベースバンド信号（図 3.10）に分類され、変調周波数を検出する BPF 処理により音波信号（図 3.11）のバイナリ情報が読み取れる。

従来の超音波通信システム

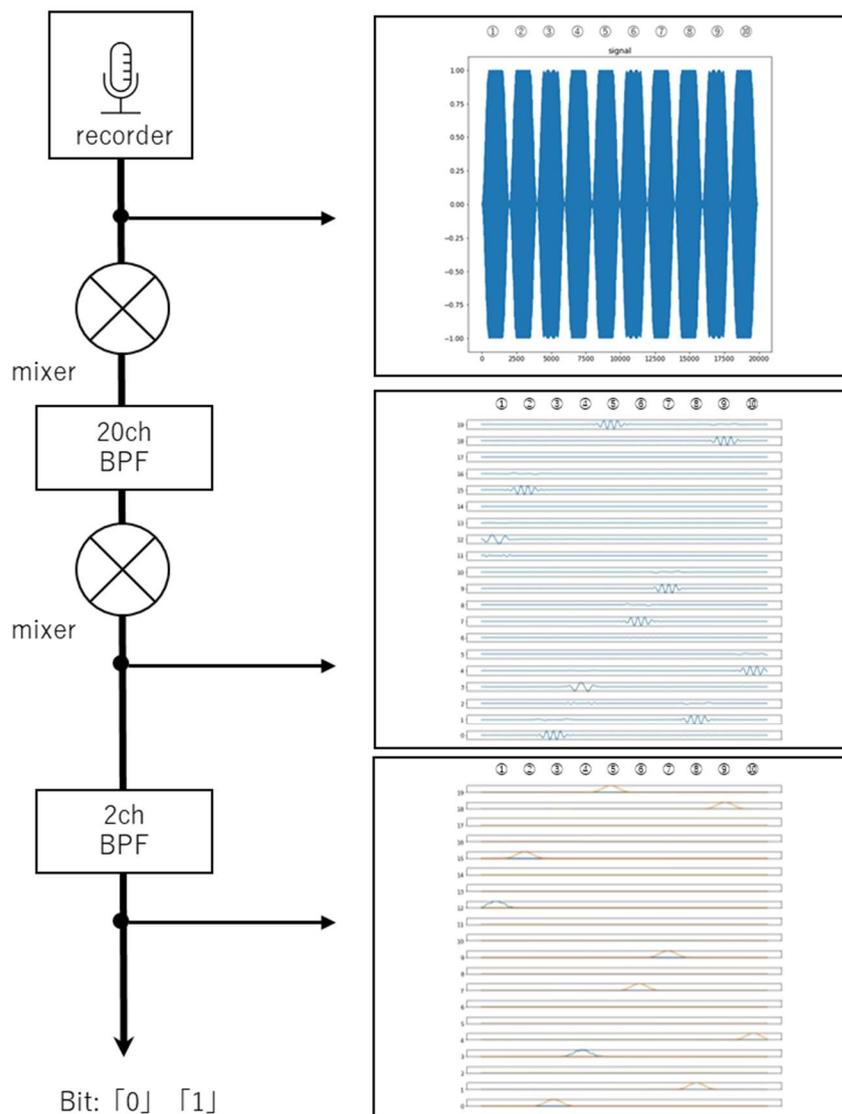


図 3.12 従来の超音波通信システムの処理の概要図

3.2.2 音波信号の配置の決定

通信において、送信信号と受信信号を照らし合わせる作業は「同期 (Synchronization)」と呼ばれる。同期は、送信側と受信側のクロックやタイミングを一致させ、信号の正確な伝送や受信を実現するための重要なプロセスである。本研究で扱う音波信号の同期方法は、400 シンボルから構成される周期的なホッピングパターンを検出ことで成り立つ。ホッピングパターンとは、あらかじめ音波信号に設定するキャリア周波数の順番であり (例えば、ホッピングパターンの 1 番目のキャリア周波数は 18.9kHz, 2 番目は 19.7kHz, , , 400 番目は 18kHz, 401 番目は 18.9kHz, , , 以降周期的に繰り返す), 録音時には、設定されているキャリア周波数のホッピングパターンと、受信した音波信号を復調処理することで得られるベースバンド信号を相関係数に用いることで同期処理が可能となる。相関係数を用いた同期方法は、信号処理や通信などの分野で広く利用される手法の一つであり、相関係数は -1 から 1 の範囲を取り、1 に近いほど正の相関があり、-1 に近いほど負の相関がある。

20ch のキャリア周波数をもつホッピングパターンの波形 $x(t)$ と、受信した音波信号のベースバンド信号 $y(t)$ の相関係数は次のように定義される。ここで、 $\text{cov}(x,y)$ は x と y の共分散、 σ_x および σ_y はそれぞれ x と y の標準偏差を示す。

$$\text{corr}(x,y) = \frac{\text{cov}(x,y)}{\sigma_x\sigma_y} \quad (3.2.1)$$

このホッピングパターンはビットエラーレート (以下, BER) の計算時に送信信号と受信信号を照らし合わせることを目的として設計されていて、20ch のキャリア周波数を組み合わせた構成により配置が決まる。図 3.13 に音波信号を構成するホッピングパターンのイメージ図を示す。なお、図 3.13 は、時間と周波数の情報を 2次元の画像として表現するスペクトログラムを表している。図 3.13 に含まれる音波信号は (A,B,C,D,E) の順にホッピングパターンが (9ch,17ch,9ch,5ch,0ch) から構成されており、バイナリ情報が (0,1,1,0,1) である。

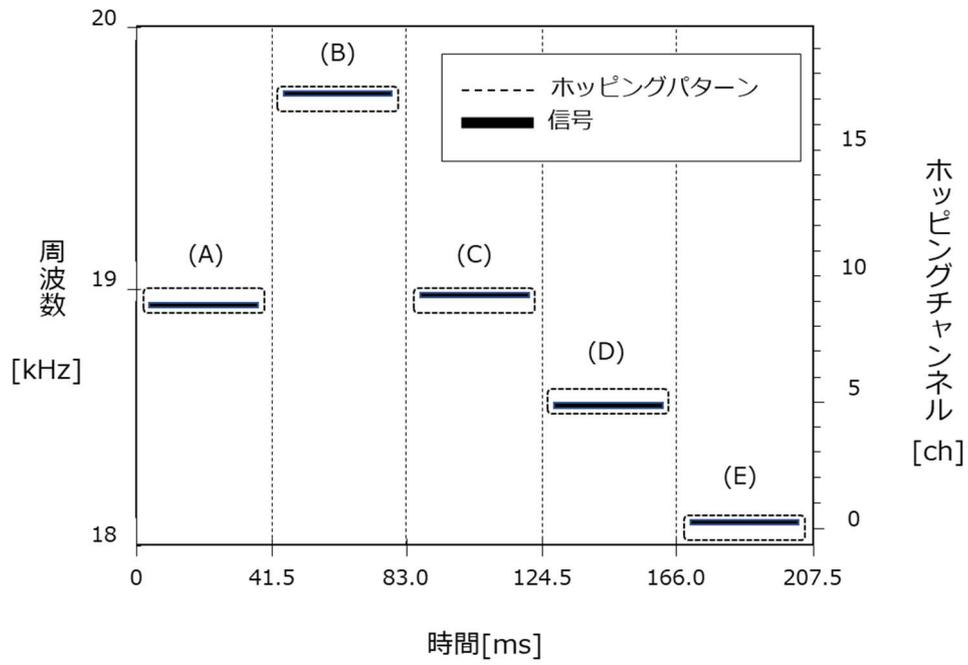


図 3.13 音波信号のスペクトログラムの時間周波数パターン

第4章

従来技術の実装

本章では、従来技術である図 3.12 の超音波通信システムを用いた際の処理結果をまとめる。

4.1 雑音特性に対する精度と課題

録音した音波信号は、400 シンボルのホッピングパターンに対して 40 シンボル分の録音データを用いて相関係数を導出し同期処理を行う。平均同期処理時間は 0.088 秒、入力した 5000 個のシンボルのバイナリ情報を計算するまでの平均処理速度は 9.68 秒であった。従来システムの識別精度を評価するため、精度評価として、5000 シンボルで構成された音波信号に対してシミュレーションにより SNR が -10dB から 10dB までとなるように白色雑音を加えた場合の、超音波通信システムの BER を図 4.1 に示す。図 4.1 は、横軸がシミュレーションにより加えた SNR[dB]を示し、縦軸が BER[%]を示しており、BER の値が 0%に近づくほど、エラー率が少なく識別精度がよいことを示している。

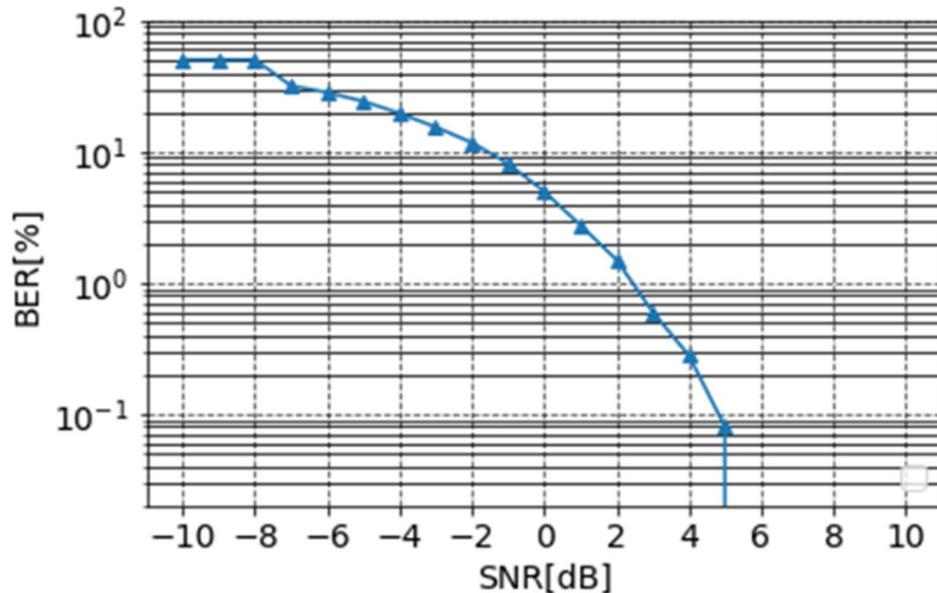


図 4.1 SNR と超音波通信システムの BER の関係

図 4.1 の結果から従来技術を用いた超音波通信システムは SNR が 6dB 以上になると BER が 0%と誤りが生じなくなることが分かる．本研究で扱う超音波通信システムは非可聴域の周波数帯を扱うため，日常で耳にするような 18kHz 以下の雑音の影響は少なく，送信する音波信号と環境の雑音レベルが一致する（SNR=0dB）ような測定状況は滅多にないため，この SNR と BER の結果から高い水準の超音波通信システムが展開できていることがわかる．また，送信端末と受信端末の通信距離が遠い場合，伝搬時に音波信号が減衰することが考えられるが SNR が 0dB 以下となる場合は，マイクで信号がほとんど聞こえていない状態であるため，同様にそのような測定状況は滅多にないと考えられる．

SNR が-8dB 以下になると BER は 50%に収束した．SNR= 10dB の復調処理により得られたベースバンド信号を図 4.2，SNR= 0dB（BER=4.98%%）を図 4.3 に示す．同様に，SNR= 10dB の BPF による抽出される各キャリア周波数の変調周波数を示す音波信号を図 4.4，SNR= 0dB を図 4.5 に示す．

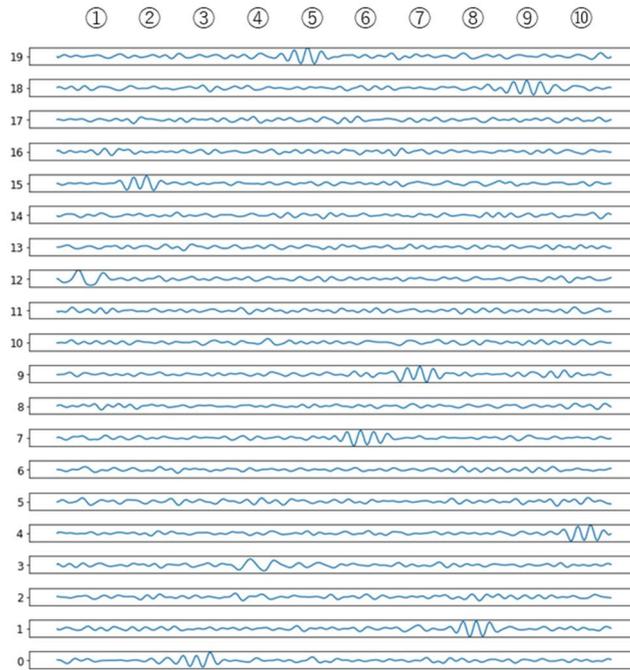


図 4.2 SNR=10dB (BER=0.00%) のベースバンド信号

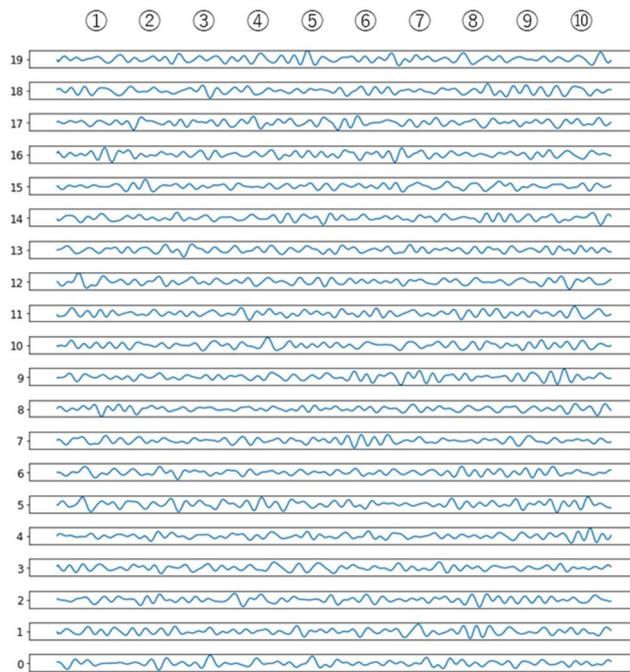


図 4.3 SNR=0dB (BER=4.98%) のベースバンド信号

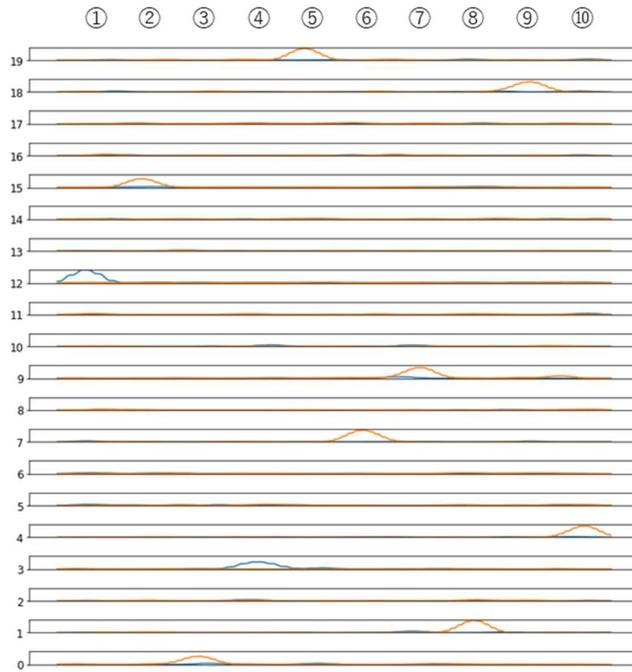


図 4.4 SNR=10dB (BER=0.00%) の変調周波数を示す音波信号

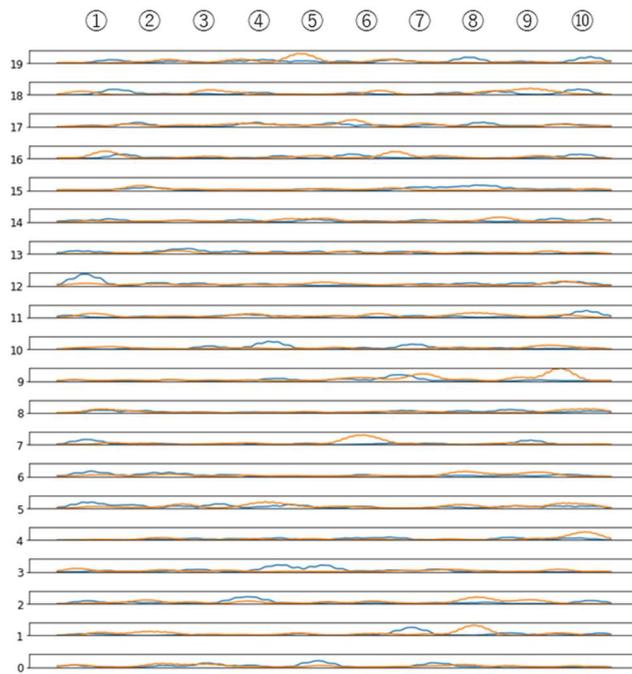


図 4.5 SNR= 0dB (BER=4.98%) の変調周波数を示す音波信号

なお、ここで示す波形は、0ch (18kHz) から 19ch(19.9kHz)のキャリア周波数から得られるベースバンド信号を示している。図 4.2, 図 4.3 の各シンボルに含まれるキャリア周波数は①19200Hz, ②19500Hz, ③18000Hz, ④18300Hz, ⑤19900Hz, ⑥18700Hz, ⑦18900Hz, ⑧18100Hz, ⑨19800Hz, ⑩18400Hz である。この復調処理後、得られたベースバンド信号は、50Hz のスペース周波数、および 100Hz のマーク周波数を抽出するように設計された BPF 処理により変調周波数が検出される。図 4.2 および図 4.4 の結果から、SNR=10dB 時には、各 BPF を通る音波信号がクリアであり、変調周波数が顕著に抽出されていることから、音波信号に含まれるバイナリ情報の識別が容易であることが分かる。一方、図 4.3 および図 4.5 の結果から、SNR=0dB 時には、各 BPF を通る音波信号に白色雑音の影響による歪がみられる。図 4.5 の変調周波数を示す音波信号からは、バイナリ情報の識別が困難であることが分かる。ただし、前述のとおり、送信する音波信号と環境の雑音レベルが一致する測定状況は滅多にないため、この問題は超音波通信システムを展開するにあたって重要な課題ではない。

4.2 周波数特性に対する精度と課題

次に周波数特性に対する精度評価として、5000 シンボルで構成された音波信号に対して白色雑音と同様にシミュレーションにより-50Hz から+50Hz までの周波数シフトを加えた場合の、従来の超音波通信システムの BER を図 4.6 に示す。図 4.6 のヒートマップは、横軸が SNR[dB]を示し、縦軸が周波数シフト [Hz]、ヒートマップの示す値が BER[%]である。BER の値が 0%に近づくほど（カラーバーの色が黒に近づくほど）、エラー率が少なく識別精度がよいことを示している。

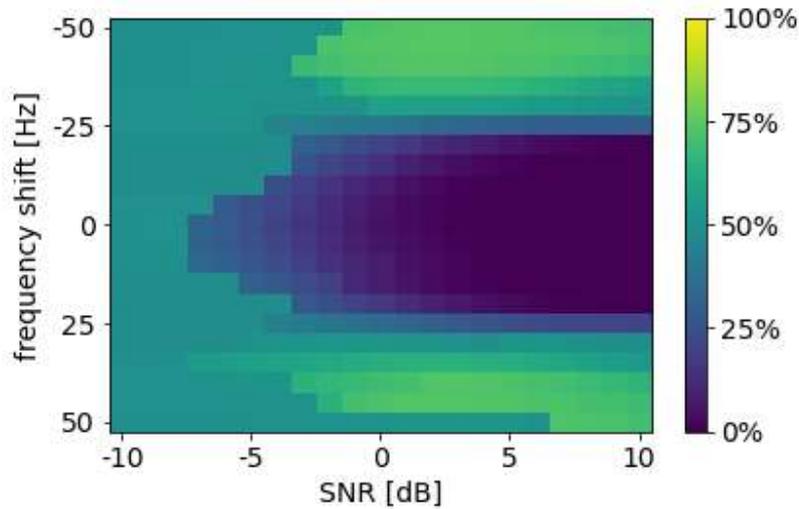


図 4.6 SNR と周波数シフトと超音波通信システムの BER の関係

従来技術を用いた超音波通信システムは図 4.1 の結果では SNR が 6dB 以上になると BER が 0% と誤りが生じたことに対して、図 4.6 の結果では、周波数シフトが 25Hz 以上発生すると 6dB 以上の SNR であっても BER が上昇していることが分かる。SNR=10dB かつ周波数シフト=25Hz の時、BER は 20.82% と大きく識別精度が低下した。この理由としては、音波信号の変調周波数を 50Hz および 100Hz と 50Hz のバンド差で設定していることが要因であり、25Hz 以上の周波数シフトは変調周波数を識別する境界線を越えることを示している。

25Hz 以上の周波数シフトが生じたときの、音波信号を確認するために、図 4.6 の結果時における SNR=10dB かつ周波数シフト=25Hz (BER=20.82%) の復調処理により得られたベースバンド信号を図 4.7 に、周波数シフト=50Hz (BER=70.24%) を図 4.8 に示す。同様に、SNR=10dB かつ周波数シフト=25Hz の BPF による抽出される各キャリア周波数の変調周波数を示す音波信号を図 4.9、周波数シフト=50Hz を図 4.10 に示す。

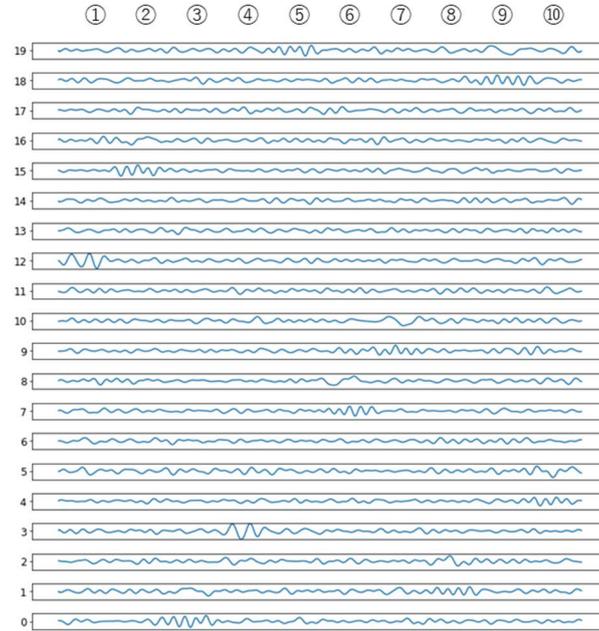


図 4.7 SNR=10dB, 周波数シフト=25Hz (BER=20.82%%) のベースバンド信号

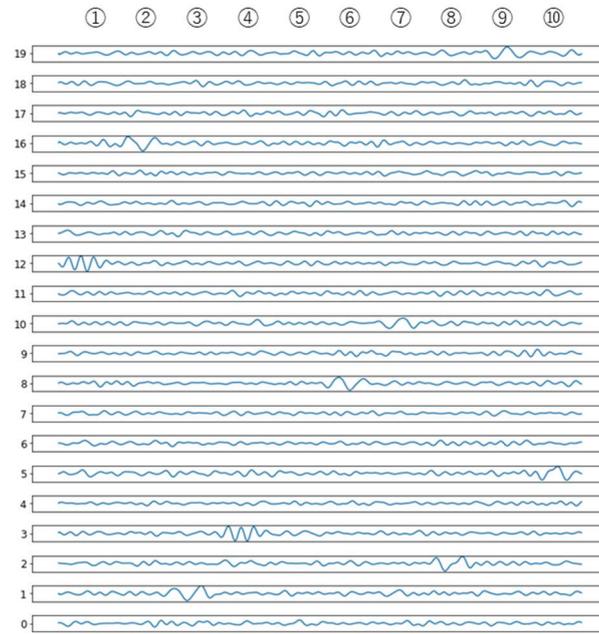


図 4.8 SNR=10dB, 周波数シフト=50Hz (BER=70.24%%) のベースバンド信号

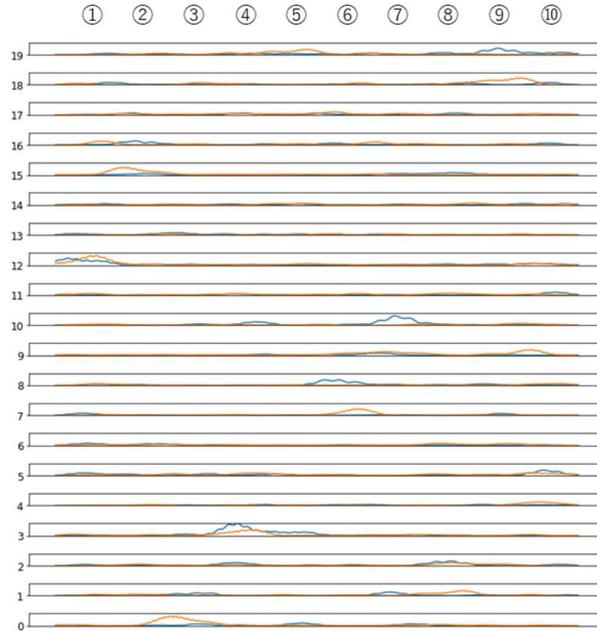


図 4.9 SNR=10, 周波数シフト=25Hz (BER=20.82%) の
変調周波数を示す音波信号

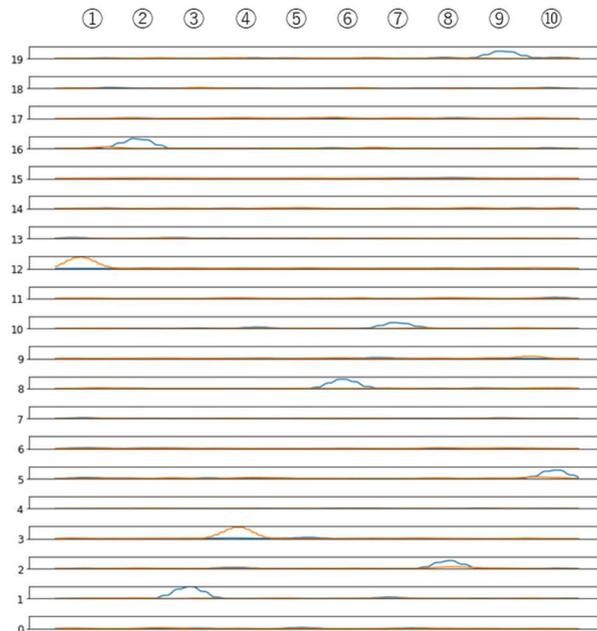


図 4.10 SNR=10, 周波数シフト=50Hz (BER=70.24%%) の
変調周波数を示す音波信号

図 4.7~図 4.10 の結果は、SNR=10dB であり雑音の影響はほとんどない。しかし、周波数シフトが 25H 発生した場合の図 4.9 の結果から、各チャンネルで得られるスペース周波数とマーク周波数が重なり、音波信号に含まれるもともとの変調周波数が、どちらの周波数であるか見分けることが困難な状態にあることが分かる。また、周波数シフトが 50H 発生した場合の図 4.10 の結果は、一見、音波信号の変調周波数がクリアに抽出されているように見えるが、図 4.2 に示す周波数シフトが発生していないときの変調周波数を示す音波信号と比較すると、もともと 50Hz のスペース周波数が 100Hz のマーク周波数と認識され、同様に 100Hz のマーク周波数をもつ音波信号が一つ上のキャリア周波数のチャンネルにてスペース周波数として抽出されていることが分かる。

このように、高い SNR を持つ音波信号であっても、周波数シフトの影響により識別精度が 20%以上も低下する問題が生じている。受信端末に周波数シフトが発生するような状況で、超音波通信システムを実用する場合（受信端末がスマートフォンだと、手を振る速度が 6m/s 程度で 30Hz 程度の周波数シフトが発生する）、使用環境によっては、通信障害が発生する問題が生じる。

この問題はどれだけ SNR がよい環境であっても発生しうる問題であり、超音波通信システムを展開するにあたって非常に重要な課題となる。

第5章

プロトタイプの開発

機械学習モデルを効果的に学習するには、大量のデータが必要であり、これには相応の計算リソースがかかる。学習プロセスの効率を向上させるために、実際の音波信号よりもシンプルで計算効率の高い簡易音波信号モデルを設計し、その学習傾向を把握するためのプロトタイプシステムを開発する。

5.1 簡易音波信号

提案するプロトタイプシステムでは、簡易音波信号を設計し現実の複雑な音波信号を再現するよりも、機械学習モデルにより基本的な特徴やパターンを抽出することに焦点を当てる。簡易音波信号は、実際の超音波通信システムで用いられる変調周波数よりも広い周波数帯域を確保し、変調周波数の差をより明確にするため 100Hz のスペース周波数と 200Hz のマーク周波数により構成する。一音(1 シンボル)の長さは 41.66ms とし、シンボル間の干渉はしないようビットレートは 24bps とする。時間と周波数の情報を 2 次元の画像として表現するスペクトログラムを利用すると、簡易音波信号は図 5.1 のように表すことができる。

この簡易音波信号に対して、学習および評価時にはシミュレーションにより周波数シフトは -30Hz から 30Hz、白色雑音は SNR が -20dB から 20dB の範囲となるように加え、データセットを生成する。なお、周波数シフトの設定値は、スマートフォンを $\pm 0.6\text{m/s}$ の速さで移動させている状況を想定して設定している。

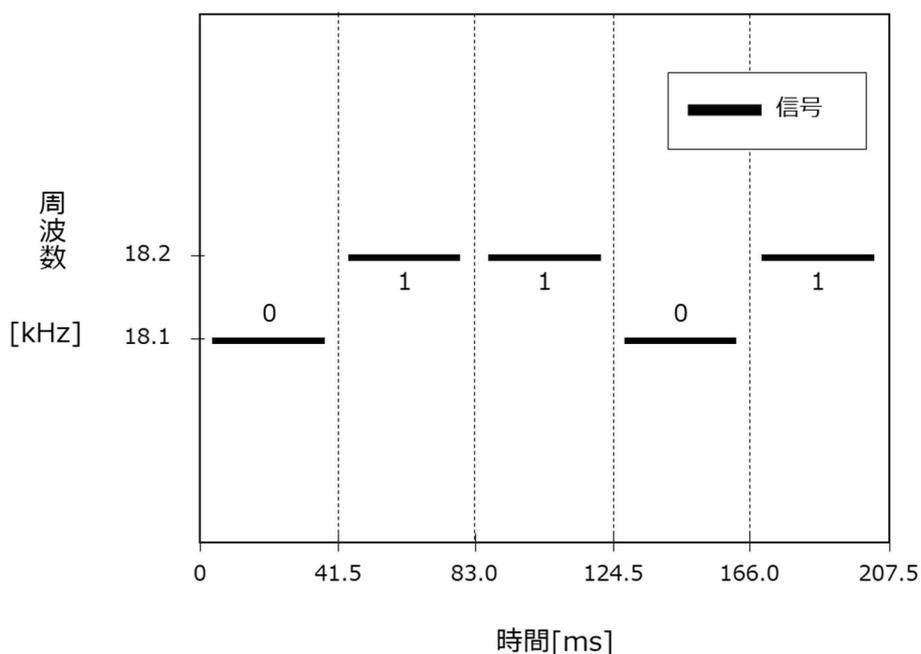


図 5.1 簡易音波信号のスペクトログラムの時間周波数パターン

5.2 前処理

提案するプロトタイプシステムは、白色雑音やインパルス性雑音、ドップラー効果や周波数シフトなどの外乱環境下から音波信号を抽出することを目的とするため、高周波成分に対して BPF や復調プロセスなどの前処理は適用しない。代わりに、短時間フーリエ変換 (short-time Fourier transform 以下, STFT) を施す。STFT により、時間変化と周波数成分の情報を 2 次元の画像として表現するスペクトログラムを用いることで、機械学習モデルによる学習を効果的に行うよう設計する。STFT で変換されるスペクトログラム $D(t, f)$ は次のように変換される。

$$D(t, f) = \int_{-\infty}^{\infty} x(\tau) \cdot \omega(\tau - t) \cdot e^{-j2\omega\tau} d\tau \quad (5.2.1)$$

$D(t, f)$ は時間 t と周波数 f の複素数をもつ STFT の結果であり, $x(\tau)$ は与えられた信号, $\omega(\tau - t)$ はハニング窓による窓関数, $\Delta\tau$ は窓関数のスライド幅である. 提案するプロトタイプシステムの STFT 処理の高速フーリエ変換 (fast Fourier transform 以下 FFT) に用いる波形の長さは 256, 窓関数のサイズは 32, 窓関数の移動幅は 2 であり, 1 シンボルごとに 1 つのスペクトログラムを作成するようにしている. また, 復調処理によりキャリア周波数をキャンセルしたベースバンド信号は, 100Hz から 200Hz までの周波数帯域で構成されるため, 学習データはベースバンド信号のスペクトログラムにおける, 0Hz から 300Hz までの周波数帯域を切り抜くことで設計した. このとき, ベースバンド信号のサンプリング周波数は 2kHz とした. 切り抜く画像サイズは 40×40 であり 1 つの画像に対して 1 シンボルとなるようにして, $[0,1]$ の二値ラベルを付与している. ラベルの値は, 1 シンボルのもつバイナリ情報を示している. 上記のパラメータを用いた STFT により得られる音波信号のスペクトログラムを図 5.2 に示す. 図 5.2 に示す簡易音波信号が有するバイナリ情報は左から (0, 1, 0, 1) である.

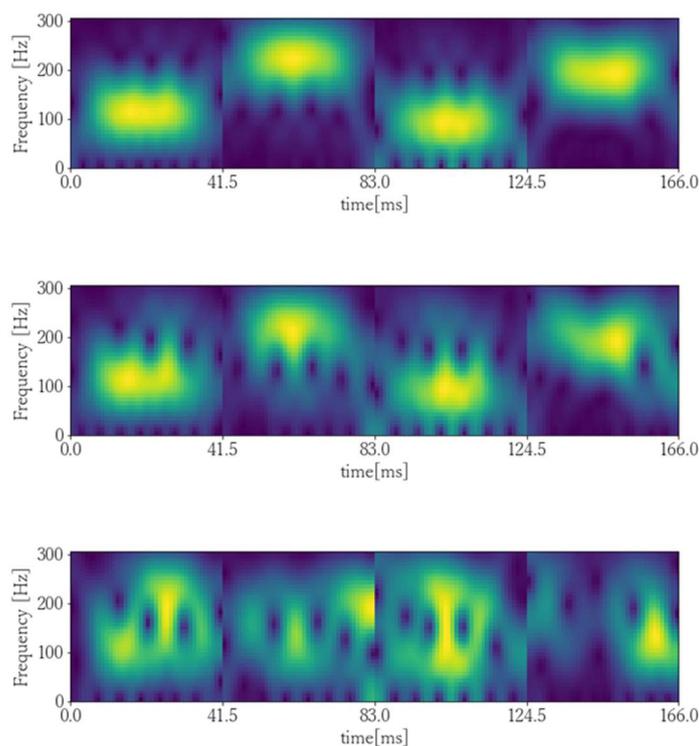


図 5.2 STFT によりスペクトログラムに変換された簡易音波信号
(上から SNR=10dB, 0dB, -10dB)

5.3 プロットタイプで設計する機械学習モデル

提案するプロットタイプシステムで設計する機械学習は、音波信号の 1 シンボルを入力したとき、そのシンボルのバイナリ情報を予測する教師あり学習を扱う。教師あり学習は、データセットにラベル（正解データ）が付いている状態で機械学習モデルを学習し、未知のデータに対する予測を行う手法である。機械学習モデルは MNIST 手書き数字データセットでよく使用される LeNet[25]を参考に CNN を設計した。MNIST は手書き数字認識のためのデータセットで、各画像は 28×28 ピクセルの小さなサイズである。LeNet は合計 2 つの畳み込み層と 2 つのプーリング層と、3 つの全結合層による構成であり、深層学習モデルと比較して層が浅く、パラメータ数も少ない小規模な機械学習モデルである。前節で説明したようにプロットタイプの学習データには 40×40 の小さな画像データを扱うため LeNet に基づいた CNN を設計した。プロットタイプで設計する CNN の構造を図 5.3 に示す。CNN は畳み込み層と Pooling 層を交互に配置し、活性化関数は ReLU を用いる。最終的に二層の dense 層と sigmoid 関数を適用して二値分類を行う。学習には ADAM[26]、損失関数として二項交差エントロピー（binary cross entropy）を用いた。

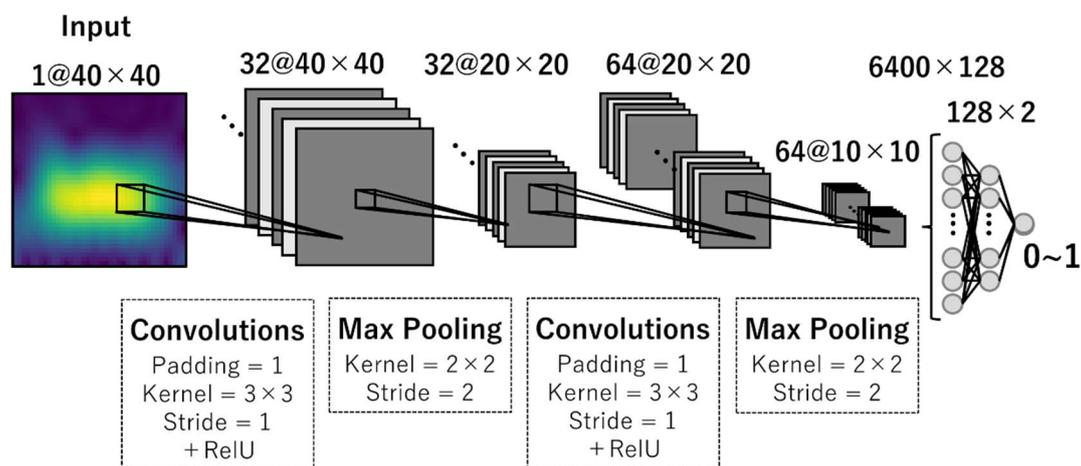


図 5.3 プロットタイプで設計する CNN 構造

5.4 プロットタイプにおける機械学習

学習の結果を、学習曲線や予測結果の比較などを用いてプロットタイプを評価する。プロットタイプの機械学習では、学習用データセットとして 200,000 個、検証用データセットとして 10,000 個用意する。各データセットは、周波数シフトは 30Hz から -30Hz、白色雑音は SNR が -20dB から 20dB の範囲となるようにシミュレーションで生成した。機械学習モデルの入力として簡易音波信号のスペクトログラムから切り抜いた 40×40 の画像を、出力として 0~1 を予測する二値分類用のデータセットを設計した。

学習率 0.001 および学習回数 (epoch 数) を 10 として学習処理を行った際の CNN の学習過程として epoch 数による Loss および Accuracy の推移を、図 5.4 および図 5.5 に示す。

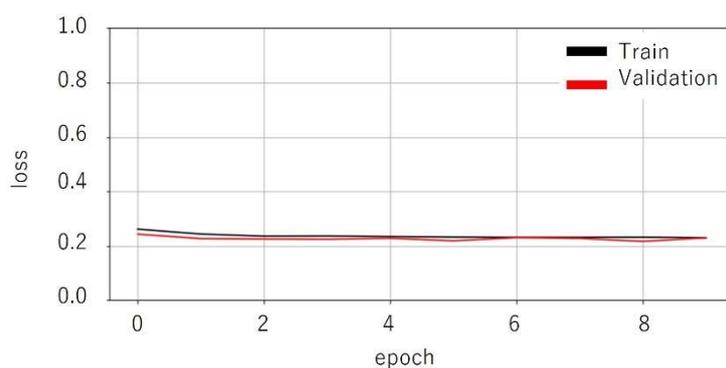


図 5.4 プロットタイプの学習過程におけるモデルの性能推移 (Loss)

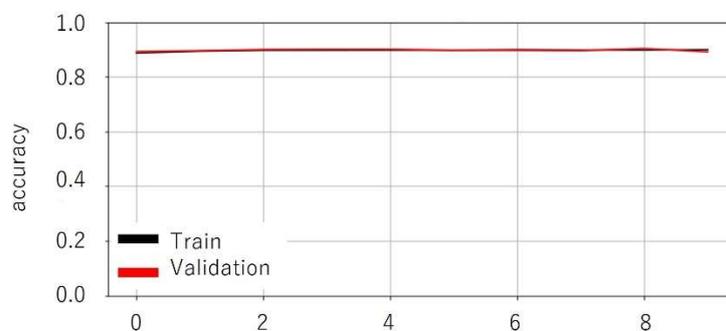


図 5.5 プロットタイプの学習過程におけるモデルの性能推移 (Accuracy)

図 5.4 および図 5.5 の学習過程において、Loss と Accuracy の曲線が示すように、プロットタイプで提案されたモデルは学習用データセットにおいて損失が減少し、同時に検証用データセットにおいても損失が適切に低減していることから、過学習が起きていないことが確認できる。このことから、パラメータ数の少ない小規模な機械学習モデルを用いた学習プロセスであっても、音波信号のスペース周波数とマーク周波数の 100Hz の差を認識することができていることが分かる。

これにより、スペクトログラムに変換された音波信号に対して、周波数の差を認識できる機械学習モデルは、超音波通信システムを設計するうえで、非常に有効的な手法であると考えられる。

5.5 簡易音波信号の録音

提案するプロットタイプシステムでは、シミュレーションによる機械学習だけでなく、実際の通信環境下における簡易音波信号の識別精度の評価を行う。ここでは実際に送信端末と受信端末を用意し二つのデバイス間で音波通信を行う。送信端末は 20kHz までの音が出せる端末であれば性能として十分であり、ノートパソコン[27]に付属しているスピーカを用いた。受信端末は、実際の活用を想定し移動可能なデバイスであるスマートフォン[28]を用い、2m の距離間で通信するように設置し 48kHz のサンプリング周波数で簡易音波信号を録音した。受信端末は音響校正器 [29]と高精度マイク[30]を用いて周波数特性の校正を行い、正確な音圧レベル (Sound Pressure Level 以下 SPL) の測定を可能とした。SPL は、空気中の音の振動が物体に加える圧力の変動の大きさでありデシベル (dB) を単位とする。

簡易音波信号は 18kHz のキャリア周波数にバイナリ情報を示す変調周波数を乗せて送信される。受信端末で録音した簡易音波信号を用いて、新たなデータセットを作成した。ここで作成したデータセットは、録音した簡易音波信号のスペクトログラムから切り抜いた 4000 個のシンボルの画像データと [0,1] の二値ラベルから構成され、以下に示す 4 つの条件の下での通信環境で、簡易音波信号を 1000 シンボルずつ録音した。

1. 受信端末の状態：固定, SNR : 5dB~10dB
2. 受信端末の状態：移動, SNR : 5dB~10dB
3. 受信端末の状態：固定, SNR : -5dB~ 0dB
4. 受信端末の状態：移動, SNR : -5dB~ 0dB

ここで、「受信端末の状態；移動」とは、通信中に±0.6m/sの速さの範囲で、受信端末を手で左右に振る、送信端末の周りで受信端末を持って歩く状態を指す。±0.6m/s という速度は±30Hzのドップラー効果が期待される値である。また、ここで示すSNRは、校正した受信端末で得られた18kHzから18.3kHzの範囲における信号の音圧レベル S_{SPL} と雑音の音圧レベル N_{SPL} の値から次のように算出される。

$$SNR = 20\log_{10} \frac{\int_{18000}^{18300} S_{SPL}(f)df}{\int_{18000}^{18300} N_{SPL}(f)df} \quad (5.5.1)$$

5.6 プロットタイプにおけるファインチューニング

転移学習の適用により，シミュレーションにより生成したデータを学習した学習済みモデルに実際に録音した音波信号データを適用する．このプロセスにより，シミュレーションデータを学習した学習済みモデルの知識を，現実の環境条件下において効果的に転移できるかを検証する．録音した簡易音波信号のデータセットから 1000 個の学習用データセット，200 個の検証用データセットを用意し，シミュレーションにより設計した学習済みモデルに対して転移学習を行う．転移学習は，一つのタスクで学習したモデルを別の関連するタスクに応用する手法であり，少量のデータセットにおいても効果的な学習を可能とする[31,32]．プロットタイプで設計した学習済みモデルは大量のシミュレーションデータを事前に学習しているため，畳み込み層では録音した音波信号の特徴を捉えることが期待できる．このように元のモデルが学習されたデータセットと，新しいタスクのデータセットが似ている場合は，転移学習の方法として最終層の Dense 層のみをパラメータ調整するファインチューニングによるパラメータ調整が効果的とされている[32]．

シミュレーションにより設計した学習済みモデルに対して学習率を 0.001 および epoch 数を 50 としてファインチューニングを行った際の CNN の学習過程として学習回数 epoch による Loss および Accuracy の推移を，図 5.6 および図 5.7 に示す．

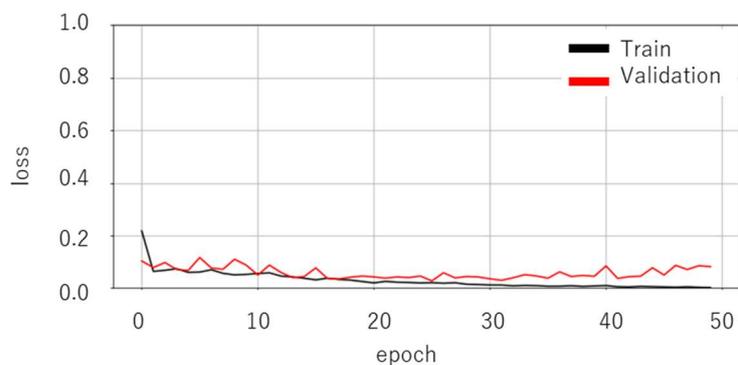


図 5.6 ファインチューニング時の学習過程におけるモデルの性能推移 (Loss)

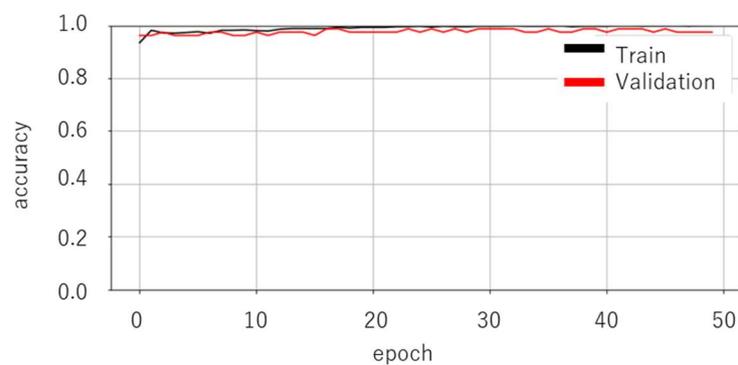


図 5.7 ファインチューニング時の学習過程におけるモデルの性能推移 (Accuracy)

図 5.6 および図 5.7 の学習過程から、対象のデータセットが少量であっても、シミュレーションによる学習済みモデルを用いた転移学習 (ファインチューニング) を行うことで、過学習を防ぎ、検証用データセットに対しても高い識別精度が得られることを確認できた。

5.7 プロットタイプの評価

プロットタイプの評価には、前節で示した 4 つの条件から学習用データセット、および検証用データセットに用いていないデータを 450 データずつ用意し評価用データとして用いる。プロットタイプの評価として、各条件の評価用データに対して、第 4 章で述べた BPF を用いた従来技術、シミュレーションにより設計した学習済みモデル、ファインチューニングにより設計した学習済みモデルにより識別を行い、そのときの BER の結果を図 5.8 に示す。

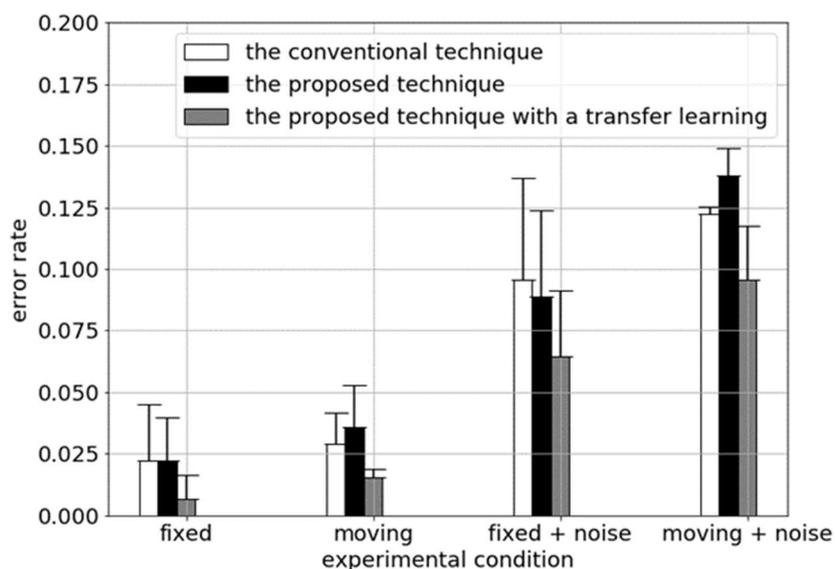


図 5.8 評価用データに対する各識別器の識別結果

まず、全体の傾向として雑音、受信端末の移動により精度は悪化することが分かる。シミュレーションにより設計した学習済みモデルは、BPFを用いた従来技術と比べて、受信端末の移動による精度の低下がみられる。この理由として、シミュレーションと録音状況によって生じる簡易音波信号の周波数成分に対する外乱特性が異なることが考えられる。シミュレーションでは周波数シフトによる外乱特性を考慮しているが、受信端末の移動による録音はドップラー効果が生じていることが分かった。シミュレーションとドップラー効果により歪む簡易音波信号の違いを図5.9に示す。1.2節に述べてように、周波数シフトは変調により信号全体の周波数が変化するが、ドップラー効果は受信端末の移動の仕方により信号の一部に周波数の変化が生じる違いがあり、その特性の違いが誤りの原因であると考えられる。一方、ファインチューニングにより設計した学習済みモデルは、実環境が与える音波信号への歪や減衰・ドップラー効果の特徴を学習することができ、すべての測定条件において識別精度が改善している。

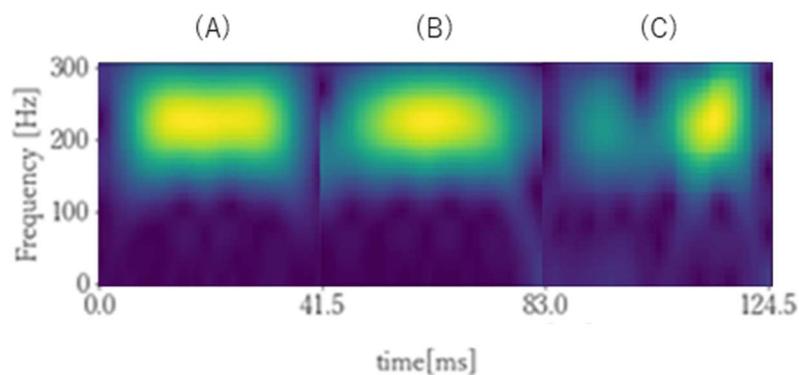


図 5.9 シミュレーションとドップラー効果により歪む簡易音波信号の違い

(A)理想の簡易音波信号, (B)周波数シフト+15Hzの簡易音波信号

(C) ドップラー効果を含む録音された簡易音波信号

5.8 プロットタイプの開発の考察・まとめ

プロットタイプシステムの設計にあたり、100Hz と 200Hz の変調周波数により構成される簡易音波信号を設計した。実際に通信を行う際は 18kHz のキャリア周波数に簡易音波信号を乗せて送信し、受信端末にて録音を行った。簡易音波信号を識別する機械学習モデルとして、LeNet を参考にした小規模な CNN モデルを設計し、シミュレーションデータの学習および録音データのファインチューニングを行った。プロットタイプの設計により小規模な機械学習モデルであっても雑音環境下における精度向上が期待できることが分かった。また、シミュレーションにより作成した学習済みモデルを元に、ファインチューニングによる転移学習を行うことで、実環境が与える音波信号への歪や減衰の特徴を捉えることができ、識別精度の向上がみられた。一方で、シミュレーションにより設計した学習済みモデルは、受信端末の移動時に精度の低下がみられた。この理由としては、シミュレーションでは周波数シフトによる外乱特性を考慮しているが、受信端末の移動による録音はドップラー効果が生じるなど、簡易音波信号の周波数成分に対する外乱特性が異なることが原因だと考えた。また、STFT により作成したスペクトログラムの周波数レンジおよびトリミングにより作成した特徴量が、想定している周波数シフトに対して不適切なサイズであることなども原因として考えられる。

プロットタイプで用いたデータは、一つ一つのシンボルに対して細かな STFT を処理し作成するため、BPF を用いた従来技術と比較して処理速度が 1000 倍以上かかる問題が生じた。本研究では設計した機械学習モデルを超音波通信システムに組み込みことを想定しており、処理速度の問題は非常に重要である。上記の理由から、一つ一つのバイナリ情報に対して予測を行う機械学習モデルは現実的ではないと考えた。

プロットタイプでは、各データから 100Hz と 200Hz の変調周波数の情報のみを埋め込みバイナリ情報の予測を行ったが、実際の音波信号はデータに含まれる信号数が増えるため、信号間の配置パターン情報などからも、雑音や周波数シフトの中から音波信号の特徴を捉えることが期待できると考えた。

第6章

提案する超音波通信システム

本章では、プロットタイプの設計で得られた特性を考慮し、実際の超音波通信システムで扱う音波信号に適応した機械学習モデルを選定し、その学習の過程と評価について説明する。

6.1 超音波通信システムの設計

プロットタイプの開発結果から、機械学習を用いることで、雑音環境および周波数シフト・ドップラー効果のような外乱を含む音波信号に対しても高い識別精度が得られることが明らかになった。しかし、処理速度の観点から、バイナリ情報に焦点を当てた機械学習アルゴリズムでは、超音波通信システムへの実装・組み込みは現実的でないことも明らかとなった。

一方、従来の超音波通信システムは、4.1.3 節にて示したように、周波数シフトが発生することで、BPF を通る音波信号が別のキャリア周波数のチャンネルをもつ音波信号に影響を与えることが問題として挙げられた。

これらの利点や課題点を考慮し、本研究では、音波信号が BPF を経由する前に、その音波信号に含まれる周波数シフトを機械学習モデルで予測し、その予測結果を従来の BPF を用いた超音波通信システムに適応させることで、周波数シフトに対応する柔軟な超音波通信システムを提案する。本研究で提案する超音波通信システムの概要図を図 6.1 に示す。

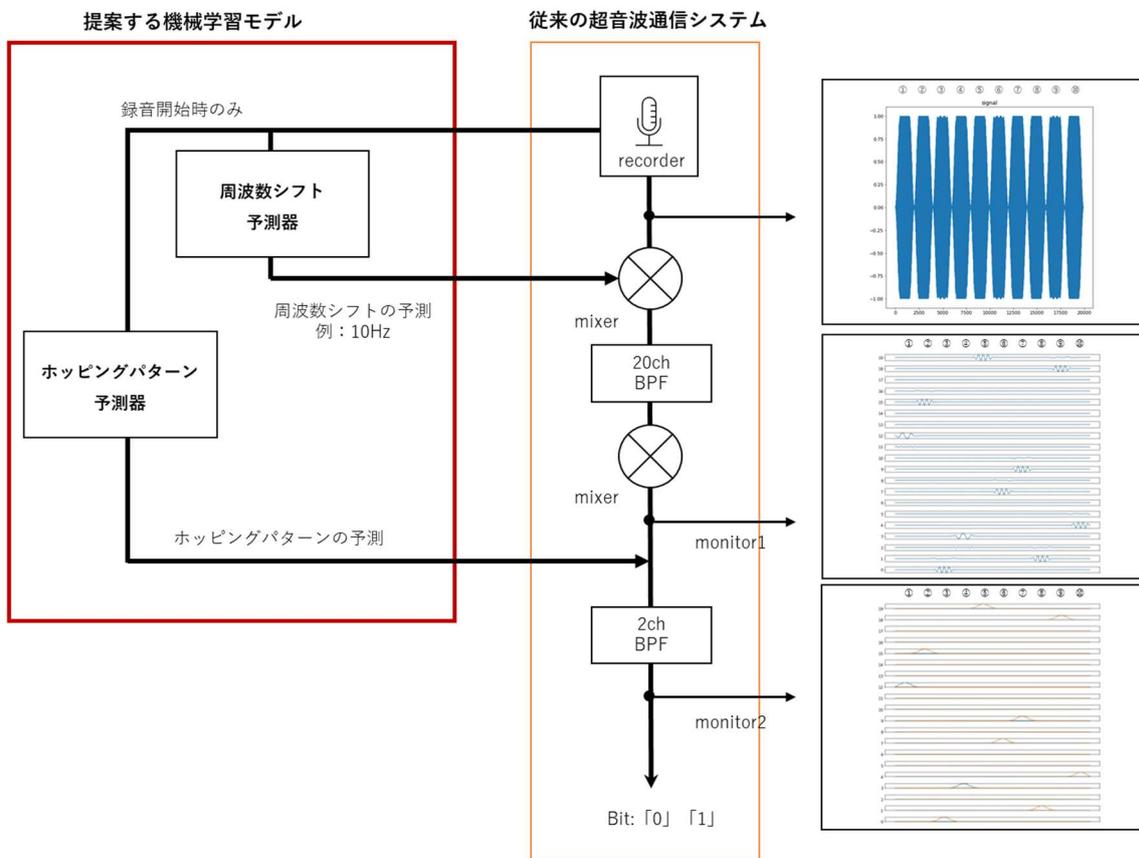


図 6.1 提案する超音波通信システムの処理の概要図

この提案手法の特色は、処理速度が遅い学習済みモデルによるタスクを、同期処理時と音波信号全体にかかる周波数シフトの予測に留め、学習済みモデルの処理回数を最小限とし、信号のバイナリ情報の予測には処理速度の速い BPF を用いた従来技術を用いることである。つまり、従来技術の周波数シフトに対する対策手段として、機械学習モデルを前処理に用いることに相当する。提案手法では、スペクトログラムに含まれる音波信号のホッピングパターンを予測し、同期処理を行う機械学習モデルと周波数シフトの値を回帰により予測する機械学習モデルの 2 つの機械学習モデルにより設計される。以後、それぞれの機械学習モデルをホッピングパターン予測器および周波数シフト予測器とする。

次に提案する機械学習モデルについて説明する。本研究で提案する機械学習モデルの概要図を図 6.2 に示す。

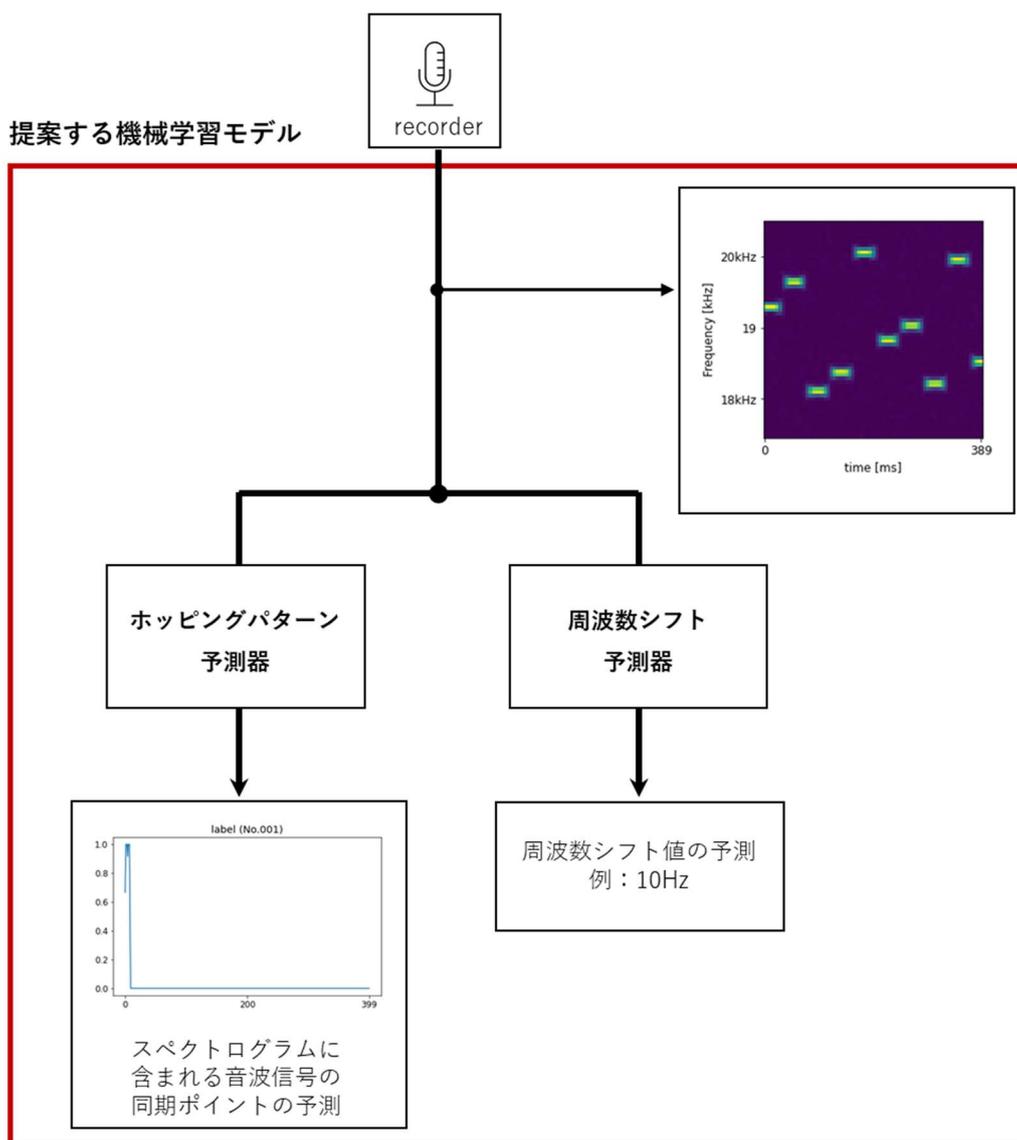


図 6.2 提案する機械学習モデルの概要図

ホッピングパターン予測器および周波数シフト予測器は、白色雑音やインパルス性雑音、ドップラー効果や周波数シフトなどの外乱環境下から音波信号の特徴を抽出することを目的とするため高周波成分に対して BPF や復調プロセスなどの前処理は適用せず、STFT によりスペクトログラムに変換された画像データを入力として扱う。変換前の音波信号を図 6.3、スペクトログラムに変換された音波信号を図 6.4 に示す。なお、それぞれの予測器で用いる STFT の FFT に用いる波形の長さおよび窓関数のサイズは 1024、窓関数の移動幅は 256 と設定している。

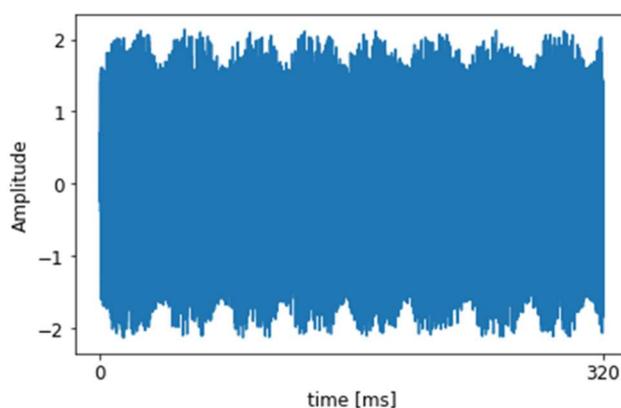


図 6.3 STFT によりスペクトログラムに変換される音波信号

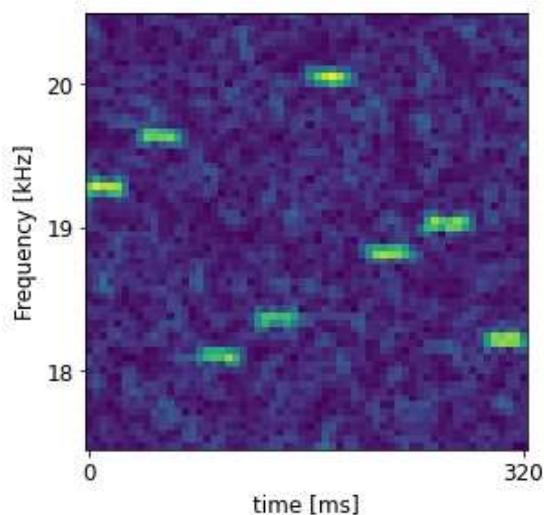


図 6.4 STFT によりスペクトログラムに変換された音波信号

各予測器は、入力されたスペクトログラムから、雑音およびドップラー効果、周波数シフトを考慮し、同期ポイントと周波数シフトの値の予測を行う。ホッピングパターン予測器は、機械学習モデルによりスペクトログラムに含まれる音波信号のホッピングパターンを分類する。分類されたホッピングパターンは、同期処理に反映させることで、雑音やドップラー効果、周波数シフトの影響を軽減することを目的として設計する。周波数シフト予測器は、機械学習モデルによりスペクトログラムに含まれる音波信号の周波数シフトの値を回帰する。回帰された周波数シフトの値を従来技術で用いる BPF の設定周波数に反映することで、周波数シフトの影響を軽減することを目的として設計する。

6.2 実験の設計

ここでは、各予測器を設計する方法について説明する。機械学習モデルは通常、大規模なパラメータを持ち、学習や推論の処理に高い計算リソースを要求する。これは高い精度を達成する一方で、実際のシステムに機械学習モデルを組み込む場合、処理速度の観点からは課題が生じる。そのため、機械学習モデルを実用するシステムに組み込む場合は処理速度向上のために簡易的なパラメータの少ない軽量なモデルを選択することが重要とされている。本研究で設計された各予測器は、実際に超音波通信システムに組み込み、実運用を想定しているため、1~5層までの軽量なモデルを設計する。本研究の機械学習アルゴリズムでは、入力に音波信号のスペクトログラムを画像と扱い、画像処理において高い識別精度が期待できる CNN・RNN・CRNN[33,34,35]の3つの機械学習モデルによる教師あり学習を扱う。実験の設計としては、先にホッピングパターン予測器に対してグリッドサーチを含む大規模な機械学習を行い、音波信号のスペクトログラムの時間周波数パターンに対して特徴を抽出しやすいモデルを最終的に選定する。グリッドサーチとは、事前に定義されたハイパーパラメータの全ての組み合わせを網羅的に試行することで最適なハイパーパラメータの組み合わせを見つけるための手法である。その後、周波数シフト予測器の学習を行う。周波数シフト予測器は、効率的に学習を行うために、ホッピングパターン予測器で最も優れた各モデルのパラメータを学習済みモデルとして、転移学習およびファインチューニングによる評価を行う。

6.2.1 ホッピングパターン予測器の設計

ホッピングパターン予測器はグリッドサーチによる大規模な機械学習を行い、CNN・RNN・CRNN の3つの機械学習モデルに対する評価を行う。ここで、RNNにはBGRU (bidirectional gated recurrent unit) [36]を用いている。以下に、本研究で行うグリッドサーチの対象パラメータを示す。

- 入力スペクトログラムのクリップサイズ： 32, 64, 128
- 畳み込み層の深さ： 1層 ~ 3層
- 畳み込み層のフィルタ数： 32, 64, 128
- 再帰層の深さ： 1層 ~ 2層
- 再帰層の隠れユニット数： 32, 64, 128
- Denseの隠れユニット数： 32, 64, 128

上記の内容でグリッドサーチを行うとCNNは81モデル分、RNNは54モデル分、CRNNは486モデル分の学習を行うことになる。畳み込み層はReLU関数、再帰層はtanh関数、各層から得られた特徴量はFlatten層を通して1次元のベクトルに変換し、Dense層とsigmoid関数による全結合層に渡るネットワークを構築した。学習にはADAM、損失関数としてbinary cross entropyを用い、最大100epoch分学習させる。学習率は0.001を初期値とし、検証用データの損失が10epochで改善されない場合、学習率が0.5ずつ減少され、最終的に学習率は0.0001となるように設定している。また、学習率を変更しても20epoch以内に精度に変化が起きない場合学習を早期終了 (Early Stopping) させる。出力層は400のユニット数を有する。これは本研究で用いる音波信号のホッピングパターンが400シンボルから構成されていることが理由である。また、400のユニットをもつ出力層にsigmoid関数を用いる理由は、ホッピングパターン予測器がマルチラベルタスクを想定しているからである。具体的には、400シンボルで構成される周期性をもつホッピングパターンの中から、入力スペクトログラムに含まれる音波信号が10~20番目のシンボルの場合、それに対応したユニットが発火するような設計をしている。例として図6.5に入力スペクトログラムと図6.6にそのラベルを示す。なお、スペクトログラムに含まれるホッピングパターンのシンボルは1を示すようにラベルを与えている。

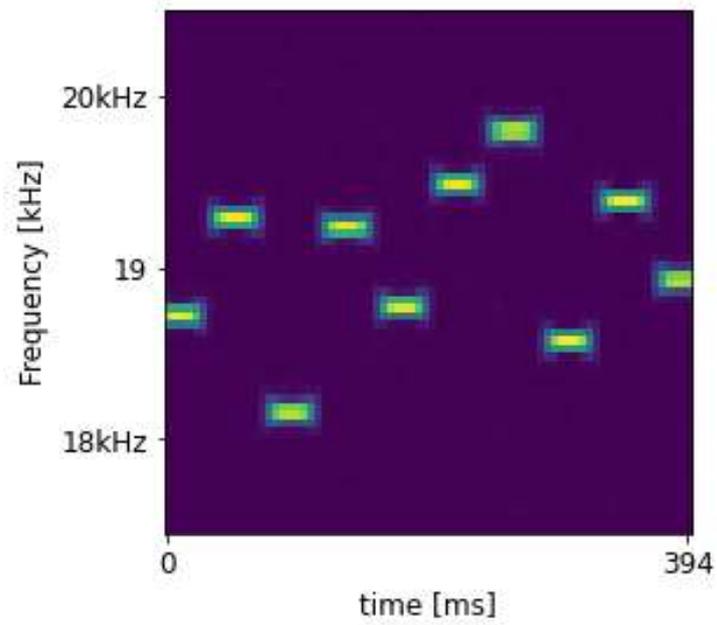


図 6.5 ホッピングパターンのうち 10~20 番目のシンボルを含む音波信号のスペクトログラム

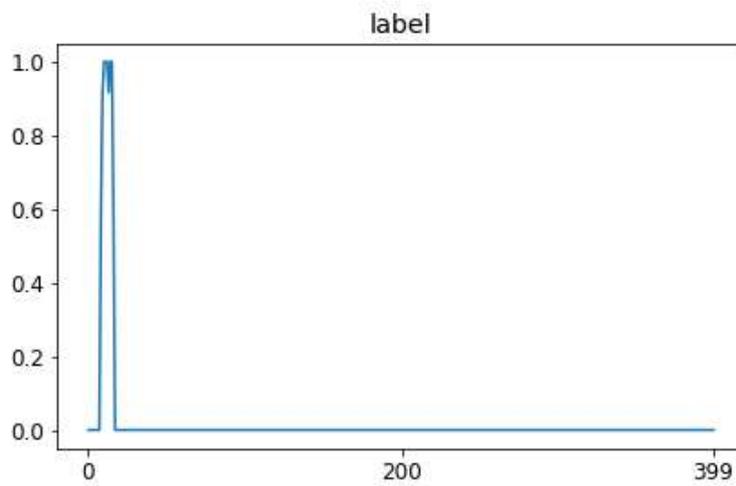


図 6.6 ホッピングパターンのうち 10~20 番目のシンボルを意味するラベル

なお、ホッピングパターン予測器の精度には再現率 (Recall) を用いる。一般的に分類タスクにおいて精度は Accuracy を計算するが、図 6.6 に示すようにホッピングパターン予測器のタスクで用いるラベルは大半の値が 0 であり、Accuracy を計算するとすべてのラベルに対して 0 と予測するだけでも高い精度となってしまう。Accuracy および Recall は次のように定義される。

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}} \quad (6.2.1)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (6.2.2)$$

それぞれの意味は以下の通りである

- True Positives (真陽性) : モデルが正例と予測し、実際にも正例であるサンプルの数
- True Negatives (真陰性) : モデルが負例と予測し、実際にも負例であるサンプルの数
- False Negatives (偽陰性) : モデルが負例と予測し、実際には正例であるサンプルの数
- Total Samples : 全体のサンプル数

6.2.2 周波数シフト予測器の設計

周波数シフト予測器はグリッドサーチによる大規模な機械学習を行わず、ホッピングパターン予測器で最もパフォーマンスが高い CNN・RNN・CRNN の3つの機械学習モデルを学習済みモデルとして、転移学習およびファインチューニングを行う。畳み込み層や再帰層、学習率や早期終了の条件はホッピングパターン予測器と同一であるが、出力層と損失関数は、回帰タスクに合わせて変更する。全結合層は Dence 層と sigmoid 関数、損失関数は平均二乗誤差 (Mean Squared Error, 以下 MSE) により構成され、出力層は1つのユニットにより 0~1 の値を回帰する。ここで、0 は周波数シフトが-100Hz, 1 は周波数シフトが100Hz である。なお、ホッピングパターン予測器の学習済みモデルに対して転移学習を行う場合は、すべてのパラメータにおいて再学習するに対して、ファインチューニングでは全結合層のパラメータのみを再学習する。そのため、周波数シフト予測器では、CNN・RNN・CRNN の3つ機械学習モデルにおいて、転移学習およびファインチューニングの合計 6 モデル分の学習を行う。周波数シフト予測器の精度には二乗平均平方根誤差 (Root Mean Squared Error, 以下 RMSE) を用いる。

6.3 データセットの収集

本研究では、機械学習の学習に使用する音波信号を取得するために、シミュレーションを活用する。シミュレーションでは、白色雑音と周波数シフトを適用し、これにより現実の環境で発生する様々な条件下での音波信号を模倣する。これにより、多様性豊かなデータセットを大量に生成することが可能となる。ホッピングパターン予測器および周波数シフト予測器の機械学習には、シミュレーションにより周波数シフトは-100Hz から+100Hz, 白色雑音は SNR が 5dB から 100dB の範囲となるように生成した簡易音波信号のスペクトログラムとそれぞれのラベルを学習用データセットとして 80,000 個、検証用データセットとして 20,000 個用意する。ただし、グリッドサーチの対象となるパラメータに入力スペクトログラムのクリップサイズがあるため、クリップサイズ (32, 64, 128) のデータセットをそ

れぞれ用意した。

シミュレーションにより加える白色雑音の SNR が 5dB~100dB と比較的高い理由は、事前の学習実験により、低い SNR を学習用データセットとして用いるよりも、比較的高い SNR を用いる方が、SNR が 0~5dB の範囲において高い識別精度が得られたからである。その事前の学習実験の結果として、図 6.7 に SNR が-5~100dB の範囲でデータセットを用意した場合、図 6.8 に SNR が 5~100dB の範囲でデータセットを用意した場合の、SNR と周波数シフトとホッピングパターン予測器の CNN モデルの Recall 結果を示す。図 6.7 および図 6.8 の SNR と周波数シフトの Recall 結果は、横軸が SNR[dB]を示し、縦軸が周波数シフト [Hz]、ヒートマップの示す値がホッピングパターンの予測結果である Recall [%]であり、Recall の値が 100%に近づくほど（カラーバーの色が黄色に近づくほど）、エラー率が少なく識別精度がよいことを示している。事前の学習実験の CNN モデルのパラメータは以下の通りである。

- 入力スペクトログラムのクリップサイズ： 128
- 畳み込み層の深さ： 3 層
- 畳み込み層のフィルタ数： 64
- Dense の隠れユニット数： 128

図 6.7 および図 6.8 の結果のように SNR が比較的高いデータセットを学習する方が、精度がよくなる傾向は、CNN・RNN・CRNN のすべての機械学習モデルに共通していた。雑音により劣化しすぎた学習データは、学習プロセスにおいて識別が困難となり誤差を生む原因となるため、提案器の学習の妨げとなっていると考えられる。

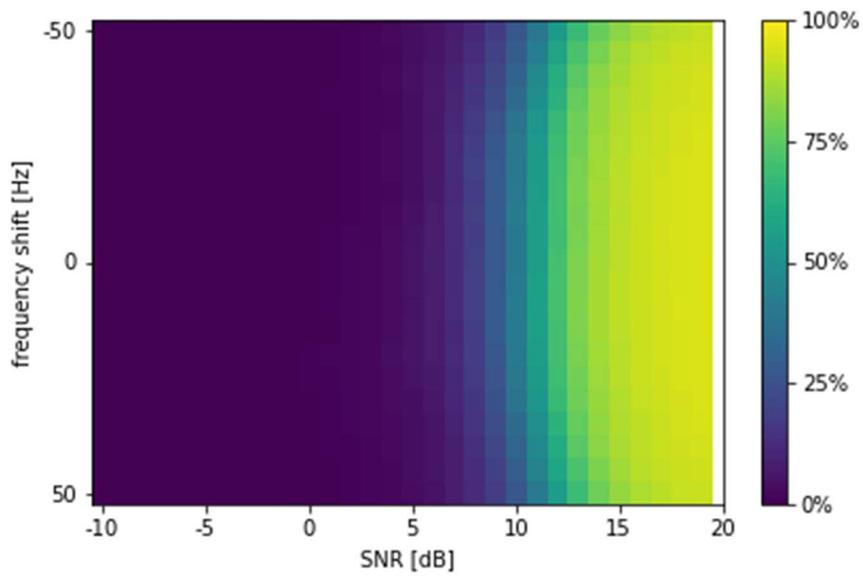


図 6.7 SNR (-5~100dB) の範囲で学習した CNN の
SNR と周波数シフトの Recall 結果

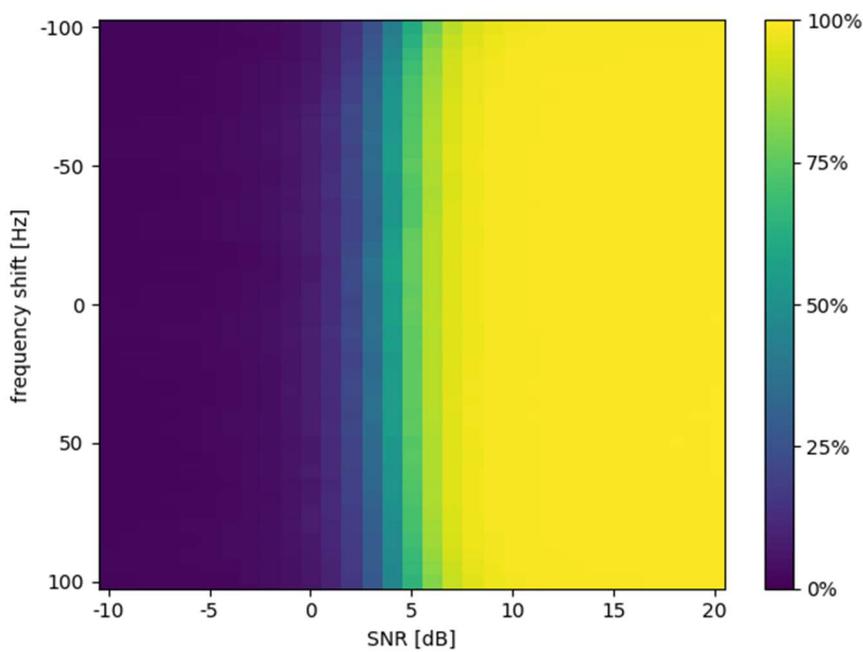


図 6.8 SNR (5~100dB) の範囲で学習した CNN の
SNR と周波数シフトの Recall 結果

6.4 提案予測器の学習および性能

ここでは、CNN・RNN・CRNN の3つの機械学習モデルに対して、シミュレーションにより作成した学習用データセットおよび検証用データセットによる各提案予測器の学習結果を示し、それぞれの学習済みモデル同士の性能評価を行う。

6.4.1 ホッピングパターン予測器の学習評価

ホッピングパターン予測器は、グリッドサーチにより CNN は 81 モデル分、RNN は 54 モデル分、CRNN は 486 モデル分の学習を行った。各モデルの学習にかかった時間を表 6.1 に示す。

表 6.1 各モデルの学習時間

	CNN	RNN	CRNN
Model数	81	54	486
計算時間[h]	8.2	2	53.6

CRNN はグリッドサーチにより調整されるパラメータ数が多いため学習時間も 2 日以上かかる結果となった。

グリッドサーチにより調整された CNN・RNN・CRNN の学習済みモデルの評価には 3 種類の評価用データセットを用いる。それぞれのデータセットは以下の特性を持つ音波信号の未知データで構成される。

1. 0Hz データセット : 周波数シフトがない場合のデータ (10000 個)
2. 100Hz データセット : 周波数シフト 100Hz のデータ (10000 個)
3. 0~50Hz データセット : 周波数シフトが 0~50Hz のデータ (50000 個)

上記の 3 種類の評価用データに対して、最も高い識別精度の平均値を得た各機械学習モデルのパラメータを表 6.2、各データセットに対する再現率 (Recall) を表 6.2 に示す。

表 6.2 グリッドサーチで得られた最適なパラメータ

	CNN	RNN	CRNN
input size	128×64×1	128×64	64×64×1
# CNN layers	3	-	1
pool size	2×2	-	1×4
# RNN layers	-	2	2
# FNN layers	1	1	1
# feature maps	64	-	32
# hidden units	-	32	64
# dense units	128	128	64
# Parameters	1,175,568	1,137,936	847,248

表 6.3 各学習済みモデルによる評価データの Recall

データセット \ モデル	CNN	RNN	CRNN
① 0Hz	96.36%	98.85%	98.85%
② 100Hz	94.73%	97.77%	98.22%
③ -50~50Hz	95.96%	98.63%	98.68%

CNN と RNN の入力スペクトログラムのクリップサイズは、パラメータの中では最も大きい 128 であり、各層の深さと Dense の隠れユニット数も最大の値となっている。一方、CRNN は畳み込み層が 1 層、再帰層が 2 層であり、CNN と RNN の総パラメータ数と比較しても小規模なモデルが、最適なパラメータとなった。表 6.3 の結果から、CRNN を用いた機械学習モデルは全てのデータセットに対して最も高い Recall が得られた。また、各機械学習モデルは周波数シフトがないデータセットに対する再現率が最も高いことが分かる。次に、表 6.2 に示す各モデルの学習過程の Loss および Recall を紹介する。図 6.9 および図 6.10 に CNN、図 6.11 および図 6.12 に RNN、図 6.13 および図 6.14 に、CRNN の Loss および Recall を示す。すべてのモデルにおいて過学習は発生しておらず、精度の良い識別が期待できる。

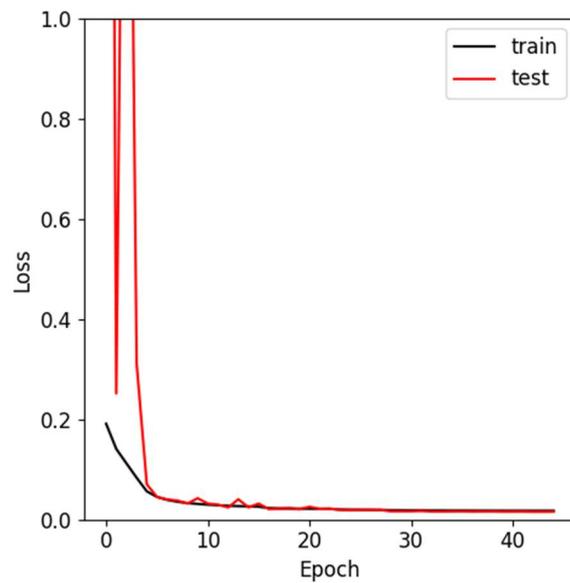


図 6.9 ホッピングパターン予測器の CNN モデルの性能推移 (Loss)

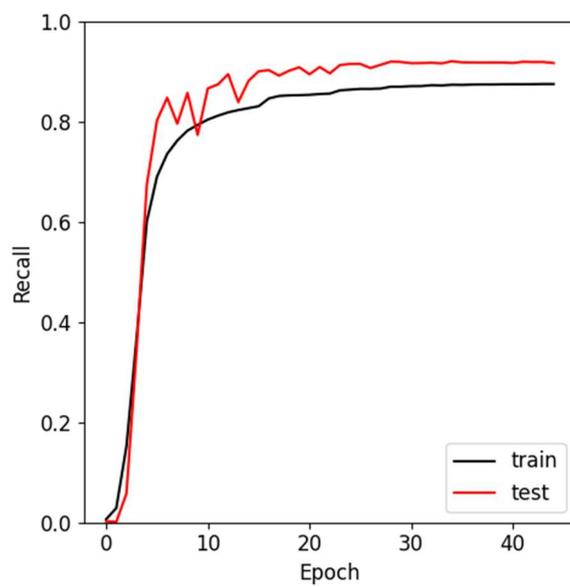


図 6.10 ホッピングパターン予測器の CNN モデルの性能推移 (Recall)

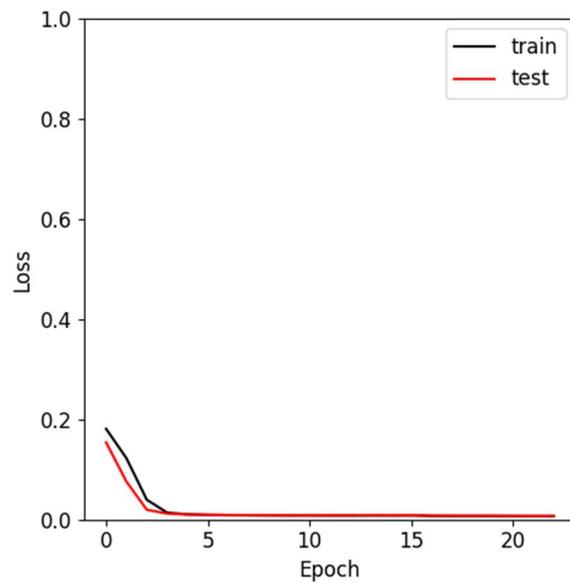


図 6.11 ホッピングパターン予測器の RNN モデルの性能推移 (Loss)

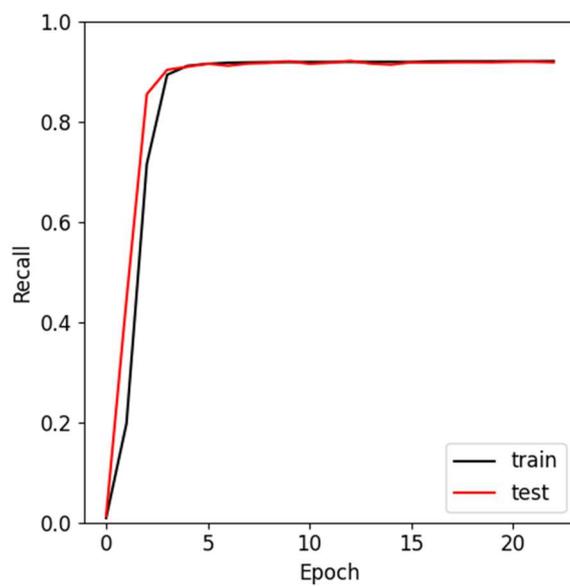


図 6.12 ホッピングパターン予測器の RNN モデルの性能推移 (Recall)

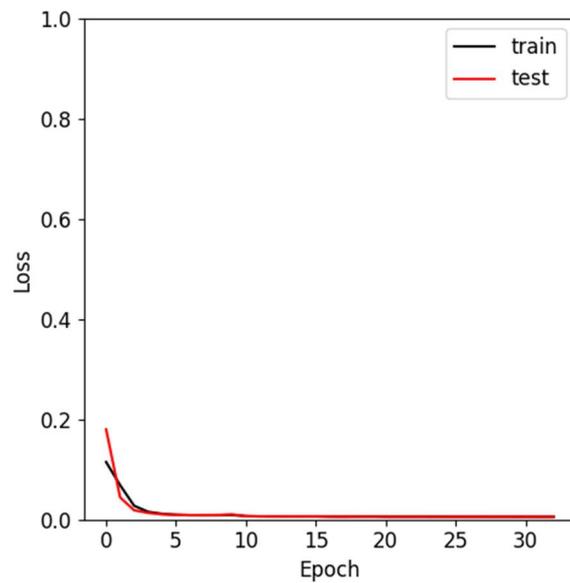


図 6.13 ホッピングパターン予測器の CRNN モデルの性能推移 (Loss)

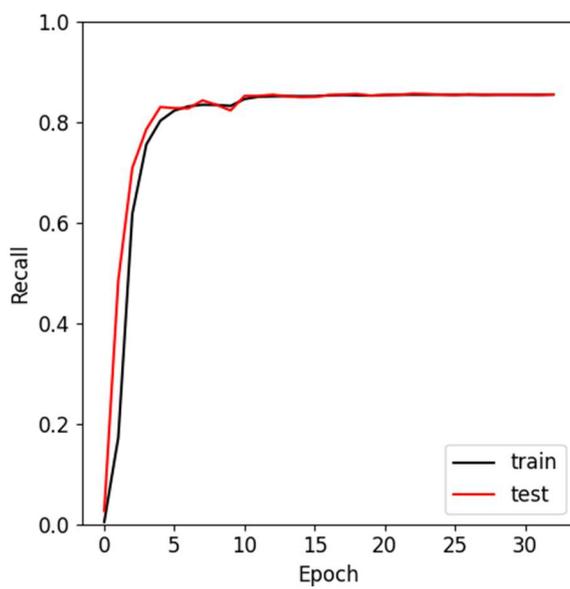


図 6.141 ホッピングパターン予測器の CRNN モデルの性能推移 (Recall)

6.4.2 ホッピングパターン予測器の性能評価

表 2 で示した CNN・RNN・CRNN の 3 つの機械学習モデルの最適なパラメータを用いてホッピングパターン予測器の性能評価を行う。

評価方法は、5000 シンボルで構成された音波信号に対して白色雑音と同様にシミュレーションにより -100Hz から +100Hz までの周波数シフトを加えた場合の評価用データセットを用意し、各モデルの Recall を計算する。SNR と周波数シフトの Recall 結果を、CNN は図 6.15、RNN は図 6.16、CRNN は図 6.17 に示す。なお、各結果の横軸は SNR[dB]、縦軸は周波数シフト [Hz]、ヒートマップは Recall [%] であり、Recall の値が 100% に近づくほど（カラーバーの色が黄色に近づくほど）、エラー率が少なく識別精度がよいことを示している。

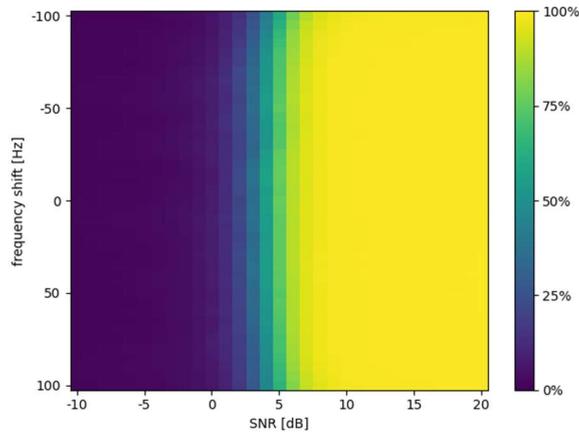


図 6.15 ホッピングパターン予測器の CNN による SNR と周波数シフトの Recall 結果

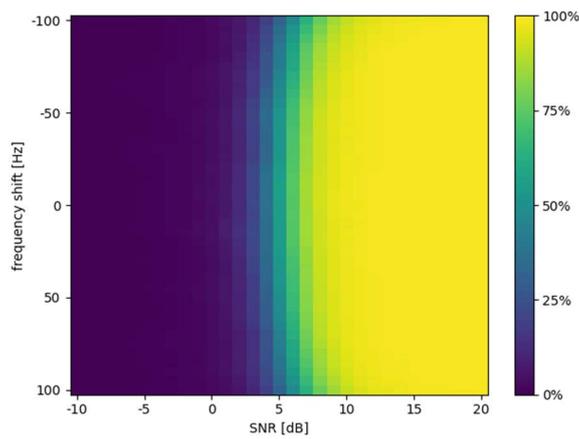


図 6.16 ホッピングパターン予測器の RNN による SNR と周波数シフトの Recall 結果

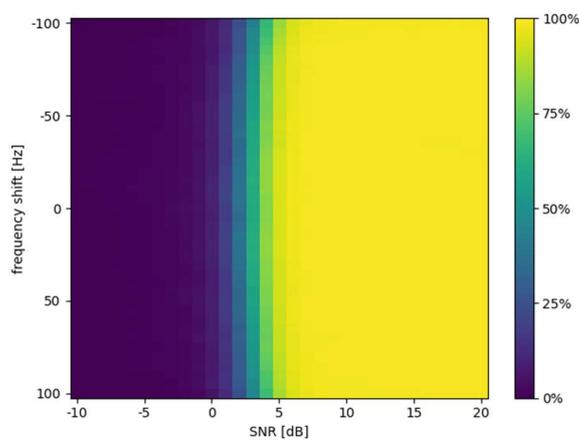


図 6.17 ホッピングパターン予測器の CRNN による SNR と周波数シフトの Recall 結果

従来技術を用いた超音波通信システムは図 4.6 の結果から、周波数シフトが 25Hz 以上発生すると 6dB 以上の SNR であっても BER が 20%以上上昇し、どれだけ SNR を高くしても周波数シフトによる BER は改善されなかった。特に、周波数シフトやドップラー効果が発生するような状況で、超音波通信システムを実用する場合（受信端末がスマートフォンだと、手を振る速度が 6m/s 程度で 30Hz 程度のドップラー効果が発生する）、使用環境によっては、通信障害が発生する問題が生じるため、非常に重要な課題であった。

一方、本研究で提案する CNN・RNN・CRNN の 3つの機械学習モデルによるホッピングパターン予測器は周波数シフトの影響に関係なくホッピングパターンの検出が行われていることが分かる。特に CRNN によるホッピングパターン予測器は、SNR が 6dB 以上あるとき、周波数シフトが 0~100Hz まで発生しても誤ることなくホッピングパターンの検出が行われた。

また、3つの機械学習モデルの中で最もスコアがよいモデルは、最もパラメータ数が少ない CRNN であった。特に CRNN は、CNN と RNN に比べると、低い SNR の音波信号に対しても精度がよい結果となった。この CRNN は畳み込み層が 1 層と再帰層が 2 層で構成されており、他のグリッドサーチを行った CRNN と比較すると RNN に近いモデルである。この 1 層の畳み込み層は、低レベルな特徴を捉えることが期待でき、近距離のシンボル間のエッジ検出や簡単な雑音除去のフィルタとしての役割が考えられる。

以上のことから機械学習を用いたホッピングパターン予測器は、周波数シフトに対して有効的な手段であることを明らかにした。ホッピングパターン予測器で設計した CNN・RNN・CRNN の 3つの機械学習モデルは周波数シフトの影響を受けることなくスペクトログラムに含まれるホッピングパターンのシンボルを予測することができ、これにより周波数シフトの有無に関係なく同期処理を行うことが可能となった。

どれだけ SNR が高い状態であっても、周波数シフトにより誤りが生じることが、従来の超音波通信システムの課題であったが、本研究で提案する機械学習モデルを用いることで、その問題を解決することができることが明らかとなった。

6.4.3 周波数シフト予測器の学習評価

周波数シフト予測器の設計では、グリッドサーチによる大規模な機械学習を行わず、ホッピングパターン予測器で最もパフォーマンスが高い CNN・RNN・CRNN の3つの機械学習モデルを学習済みモデルとして、転移学習およびファインチューニングを行う。

図 45 および図 46 に CNN，図 47～図 48 に RNN，図 49～図 50 に，CRNN に対して転移学習を行ったときの Loss および RMSE を示す。すべてのモデルにおいて過学習は発生しておらず，精度の良い識別が期待できる。

同様に，図 51 および図 52 に CNN，図 53～図 54 に RNN，図 55～図 56 に，CRNN に対してファインチューニングを行ったときの Loss および RMSE を示す。

転移学習およびファインチューニングの学習過程を比較すると，すべてのモデルにおいて過学習は発生しておらず，精度の良い識別が期待できる。また，転移学習の方が，精度よく RMSE が収束していることが分かる。

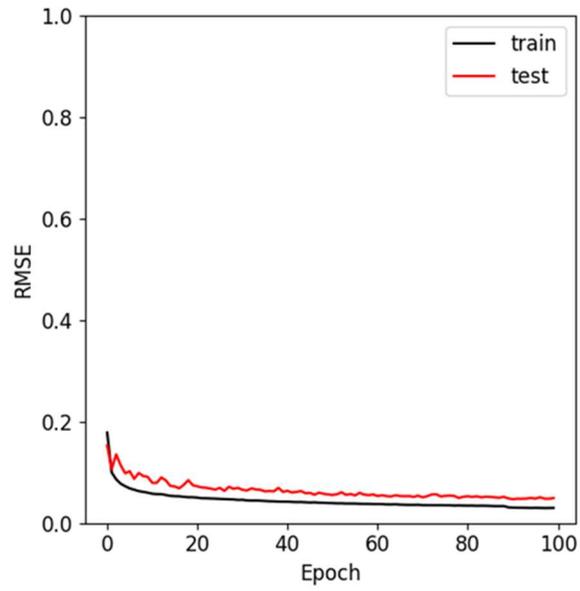


図 6.182 転移学習による周波数シフト予測器の CNN モデルの性能推移 (Loss)

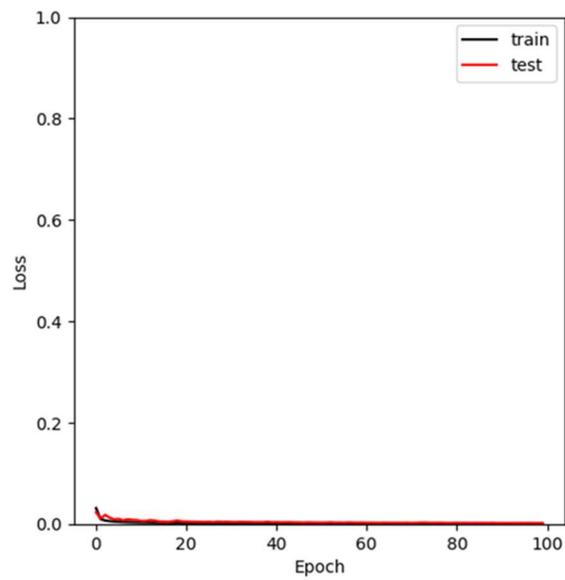


図 6.19 転移学習による周波数シフト予測器の CNN モデルの性能推移 (RMSE)

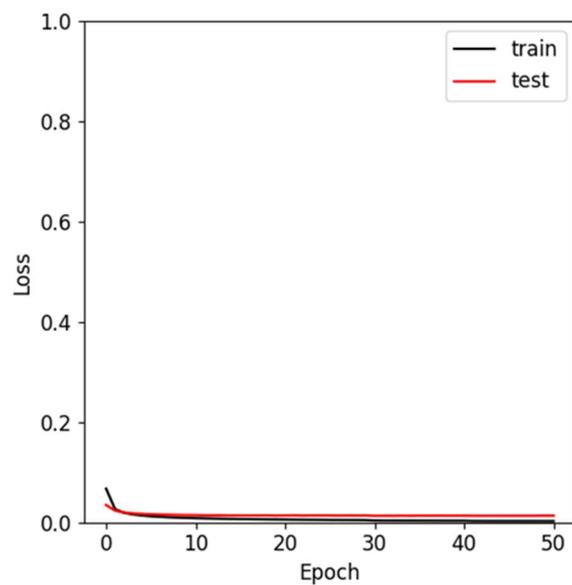


図 6.20 転移学習による周波数シフト予測器の RNN モデルの性能推移 (Loss)

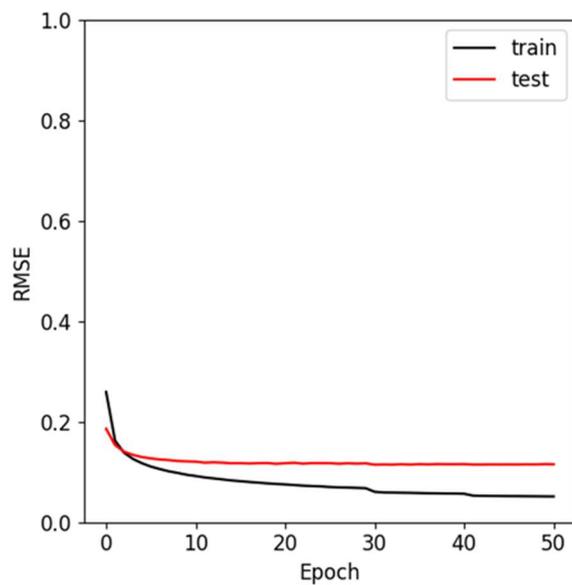


図 6.21 転移学習による周波数シフト予測器の RNN モデルの性能推移 (RMSE)

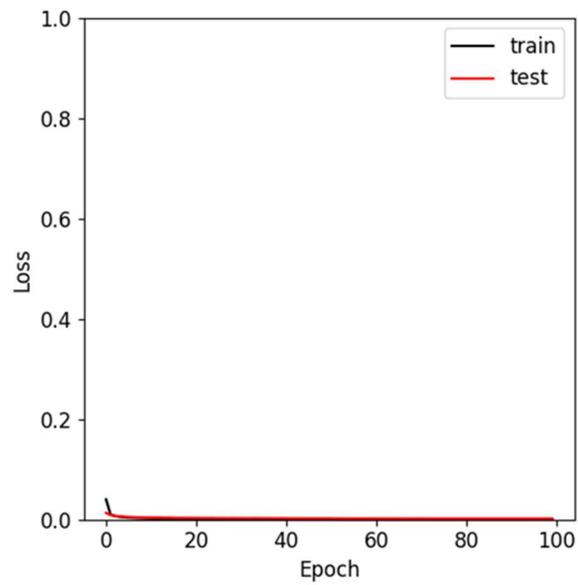


図 6.22 転移学習による周波数シフト予測器の CRNN モデルの性能推移 (Loss)

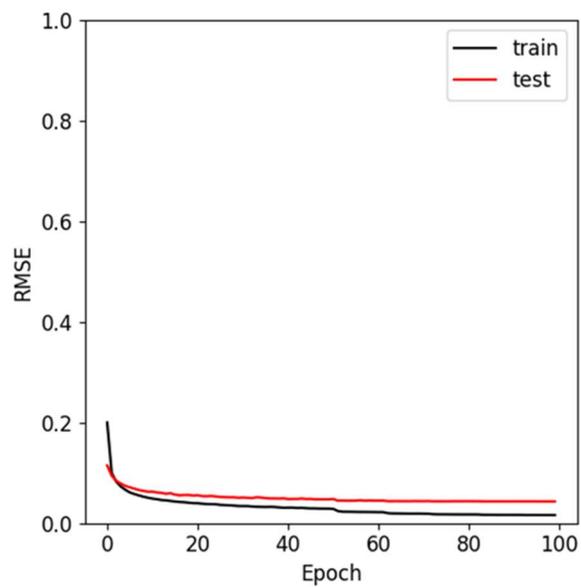


図 6.23 転移学習による周波数シフト予測器の CRNN モデルの性能推移 (RMSE)

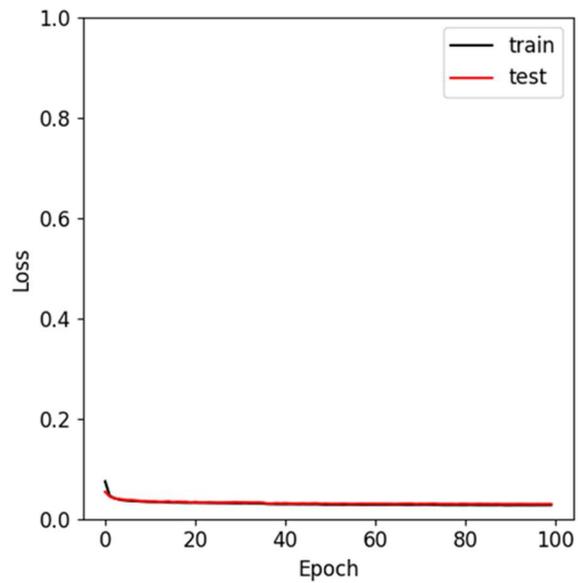


図 6.24 ファインチューニングによる周波数シフト予測器の CNN モデルの性能推移 (Loss)

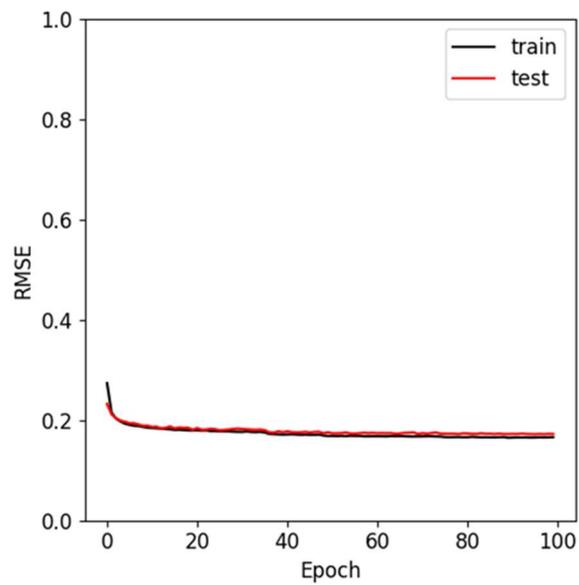


図 6.25 ファインチューニングによる周波数シフト予測器の CNN モデルの性能推移 (RMSE)

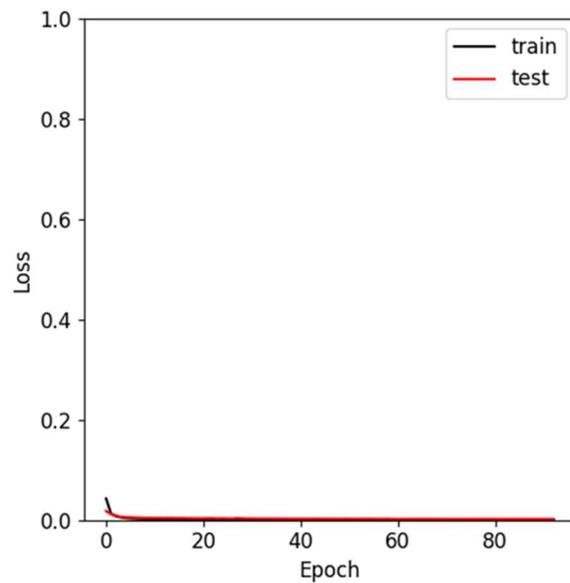


図 6.26 ファインチューニングによる周波数シフト予測器の RNN モデルの性能推移 (Loss)

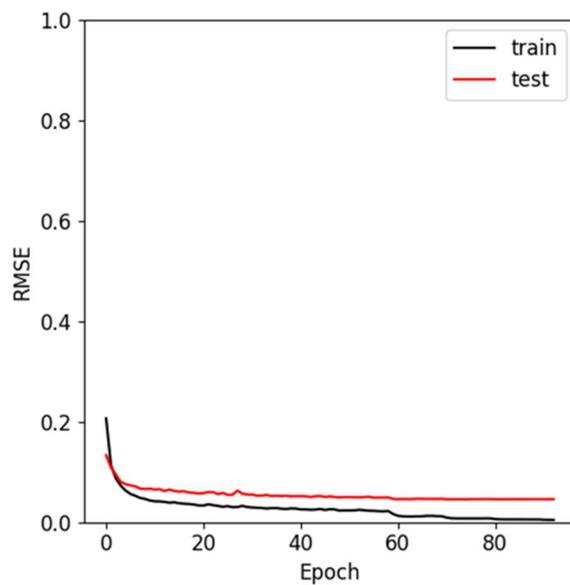


図 6.27 ファインチューニングによる周波数シフト予測器の RNN モデルの性能推移 (RMSE)

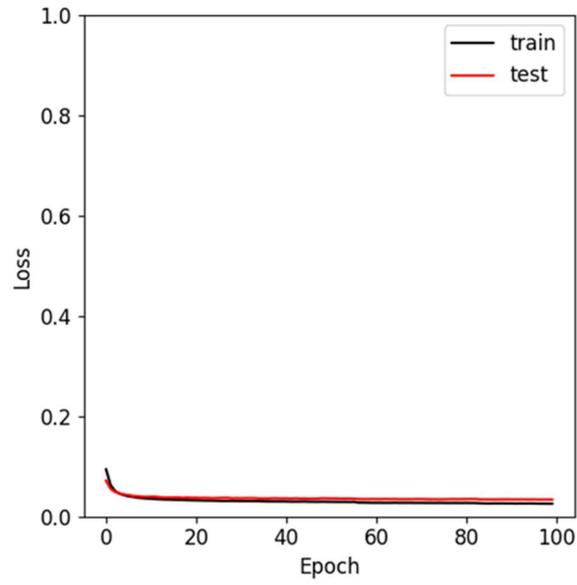


図 6.28 ファインチューニングによる周波数シフト予測器の CRNN モデルの性能推移 (Loss)

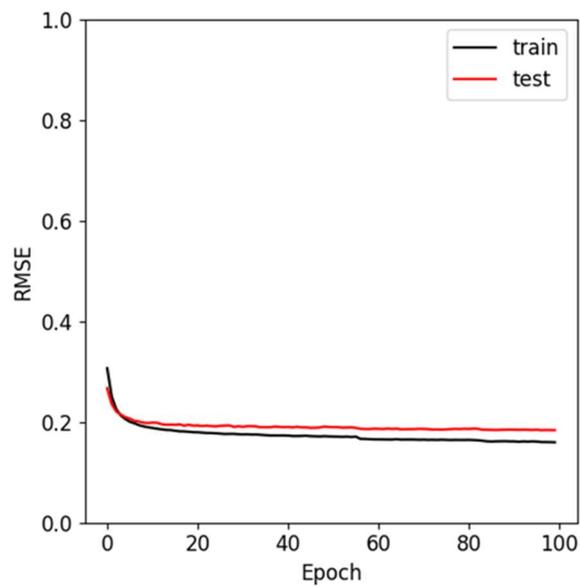


図 6.29 ファインチューニングによる周波数シフト予測器の CRNN モデルの性能推移 (RMSE)

6.4.4 周波数シフト予測器の性能評価

ここでは、ホッピングパターン予測器で最もパフォーマンスが高い CNN・RNN・CRNN の3つの機械学習モデルを学習済みモデルとして、転移学習およびファインチューニングを行ったときの周波数シフト予測器の性能評価を行う。

評価方法は、5000 シンボルで構成された音波信号に対して白色雑音と同様にシミュレーションにより-100Hz から+100Hz までの周波数シフトを加えた場合の評価用データセットを用意し、各モデルで周波数シフトを予測しその精度を計算する。各結果の横軸は SNR[dB]、縦軸は周波数シフト[Hz]、ヒートマップは RMSE [%]であり、RMSE の値が 0%に近づくほど（カラーバーの色が黄色に近づくほど）、エラー率が少なく識別精度がよいことを示している。図 6.30 に CNN、図 6.31 に RNN、図 6.32 に CRNN に対して転移学習を行ったときの周波数シフト予測器による SNR と周波数シフトの精度結果を示す。同様に、図 6.33 に CNN、図 6.34 に RNN、図 6.35 に CRNN に対してファインチューニングを行ったときの周波数シフト予測器による SNR と周波数シフトの精度結果を示す。

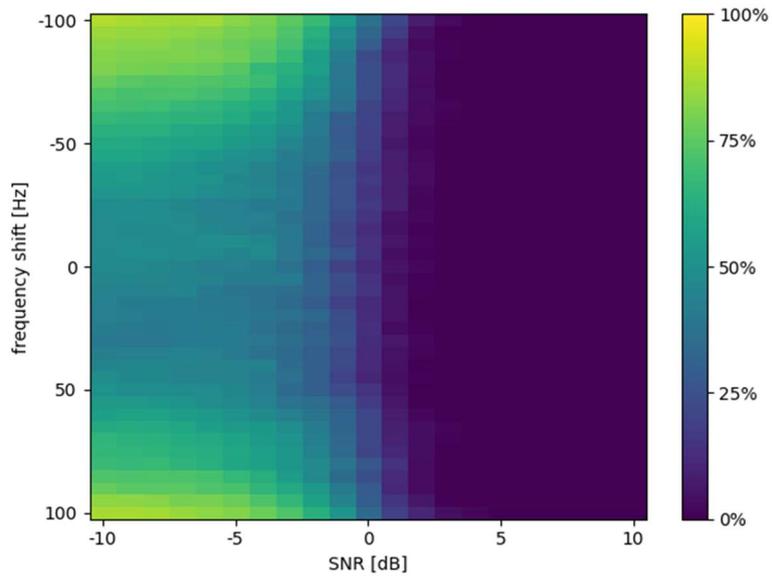


図 6.30 周波数シフト予測器の CNN の転移学習による SNR と周波数シフトの RMSE 結果

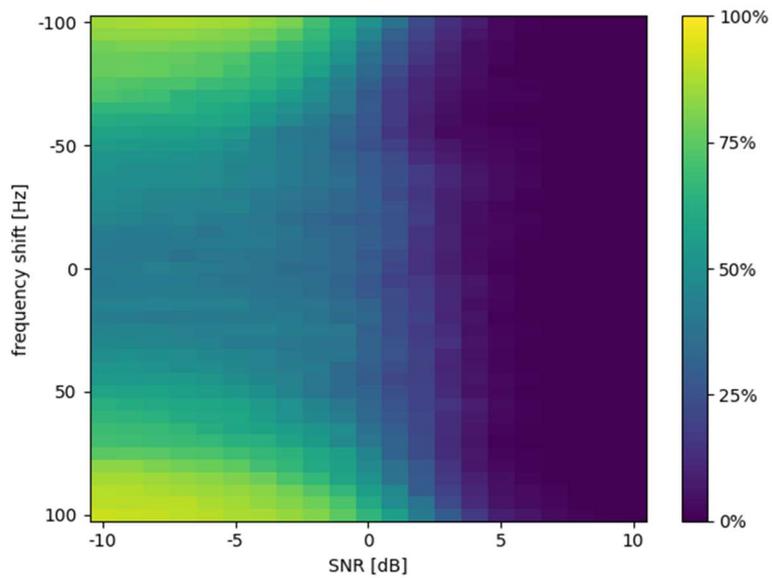


図 6.31 周波数シフト予測器の RNN の転移学習による SNR と周波数シフトの RMSE 結果

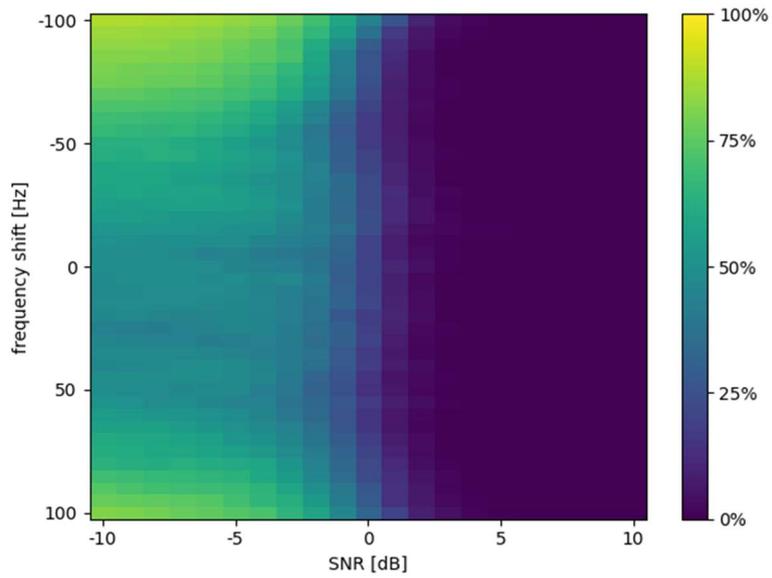


図 6.32 周波数シフト予測器の CRNN の転移学習による SNR と周波数シフトの RMSE 結果

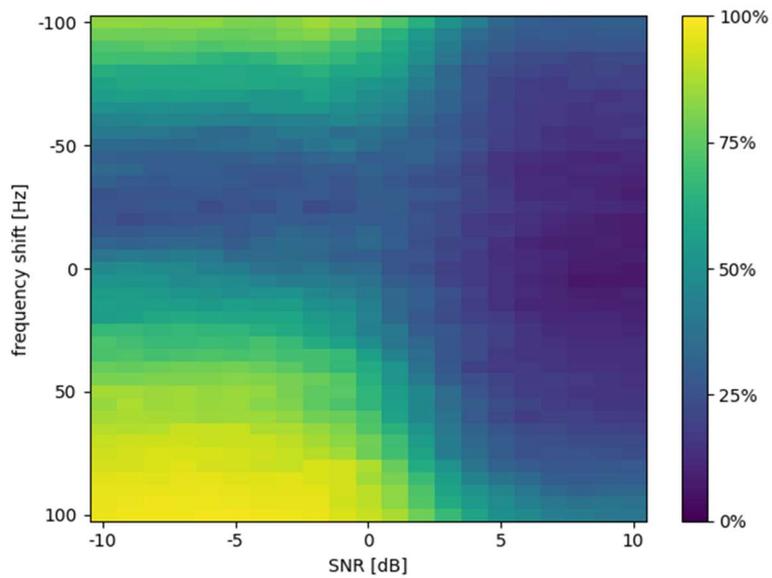


図 6.33 周波数シフト予測器の CNN のファインチューニングによる SNR と周波数シフトの RMSE 結果

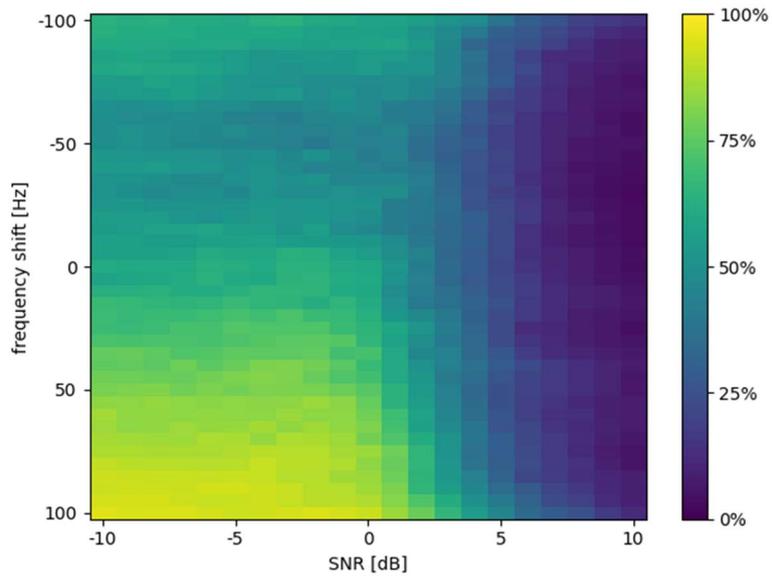


図 6.34 周波数シフト予測器の RNN のファインチューニングによる
SNR と周波数シフトの RMSE 結果

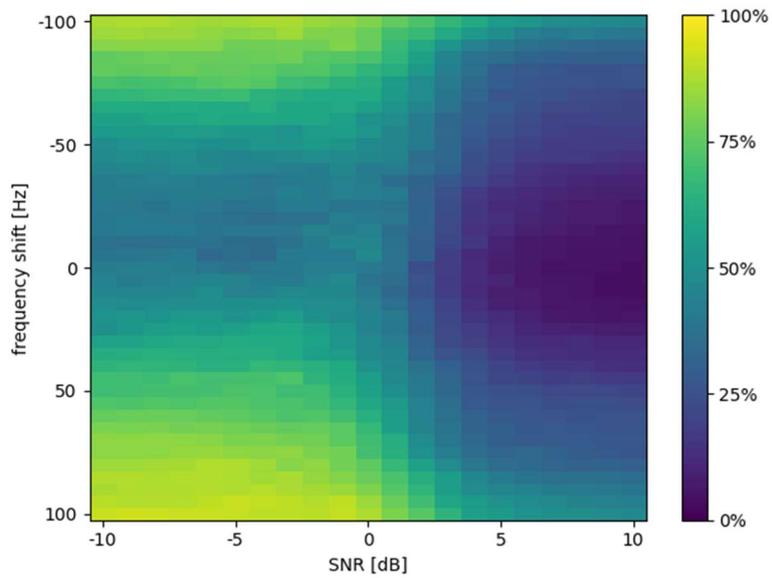


図 6.35 周波数シフト予測器の CRNN のファインチューニングによる
SNR と周波数シフトの RMSE 結果

性能評価の基準として、従来技術の超音波通信システムでは、25Hz以下の周波数シフトかつSNRが6dB以上のとき、BERは0%であったため、周波数シフト予測器により予測される周波数シフトの値は±20%までを許容範囲として評価する。

図 6.36 では、代表として CRNN による周波数シフト予測器の性能評価を行う。

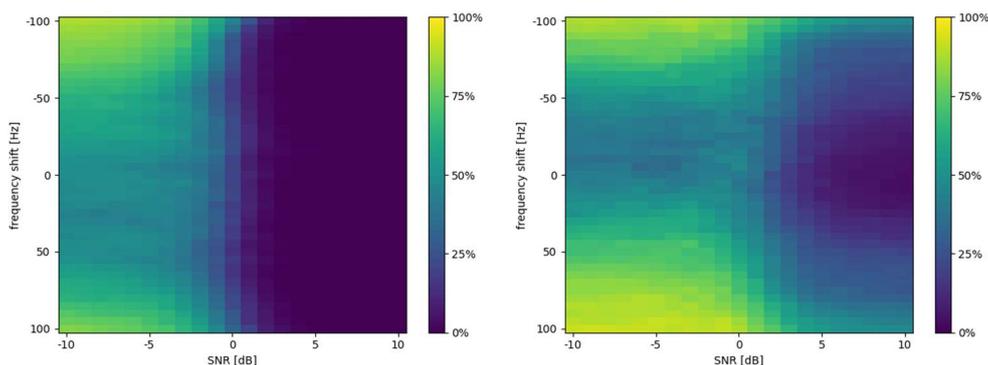


図 6.36 周波数シフト予測器の CRNN の転移学習とファインチューニングによる SNR と周波数シフトの RMSE 結果 (左: 転移学習, 右: ファインチューニング)

まず転移学習およびファインチューニングを行ったときの周波数シフト予測器の性能は明らかに転移学習により再学習を行った予測器が、精度よく周波数シフトおよび SNR の特性を学習していることが分かる。ホッピングパターンのマルチラベルタスク時に各機械学習モデルが抽出していた特徴量と、周波数シフトの値を回帰タスクとして予測際に、抽出される特徴量は、ファインチューニングが成り立っていることから似てはいるが、転移学習の方が良い結果になっていることから異なる特徴量を抽出しているということが考えられる。

以上から、録音した音波信号を転移学習またはファインチューニングをする際は深い層のパラメータ調整が重要となると考えられる。以後の周波数シフト予測器による実験では、転移学習により再学習を行った CNN・RNN・CRNN の 3 つの機械学習モデルを用いる。

6.5 性能評価

ここでは、図 3.12 に示す従来の超音波通信システムの処理方法と、図 6.1 に示す提案する超音波通信システムの処理方法を用いて、評価用データセットの BER を計算し、雑音および周波数シフトを含む音波信号に対する精度評価を行う。本研究では、CNN・RNN・CRNN の 3 つの機械学習モデルを用いて、ホッピングパターン予測器および周波数シフト予測器を設計した。そこで、提案する超音波通信システムの性能評価では各機械学習モデルで設計された提案予測器により評価を行う。ここでは、2 つの評価用データセットを用意して、従来の超音波通信システムと提案する超音波通信システムの精度比較を行う。

評価用データセット①：

5000 シンボルで構成された音波信号に対してシミュレーションにより SNR が-10dB から 10dB までとなるように白色雑音を加えたデータセット

評価用データセット②：

5000 シンボルで構成された音波信号に対してシミュレーションにより SNR が-10dB から 10dB までとなるように白色雑音を、かつ-50Hz から+50Hz までの周波数シフトを加えたデータセット

それぞれのデータセットに対する超音波通信システムの BER と SNR の関係を図 6.37、図 6.38 に示す。また、周波数シフトによる BER の関係を示すため、図 6.39 に従来の超音波通信システム、図 6.40 に CNN、図 6.41 に RNN、図 6.42 に CRNN の提案予測器を用いた超音波通信システムによる BER の精度を示すヒートマップを示す。なお、図 6.40-6.42 のヒートマップは、横軸が SNR[dB]を示し、縦軸が周波数シフト[Hz]、ヒートマップの示す値が BER[%]である。BER の値が 0%に近づくほど(カラーバーの色が黒に近づくほど)、エラー率が少なく識別精度がよいことを示している。

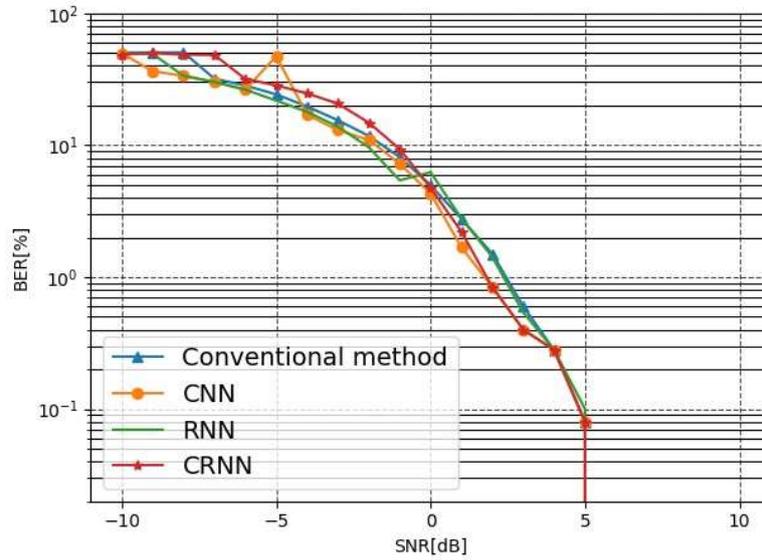


図 6.37 各超音波通信システムの雑音特性に対する精度 (評価用データセット①)

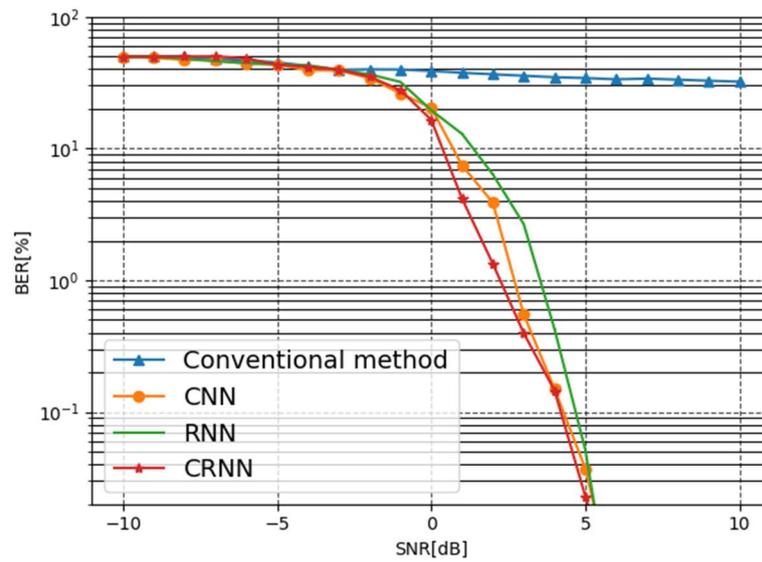


図 6.38 各超音波通信システムの雑音特性に対する精度 (評価用データセット②)

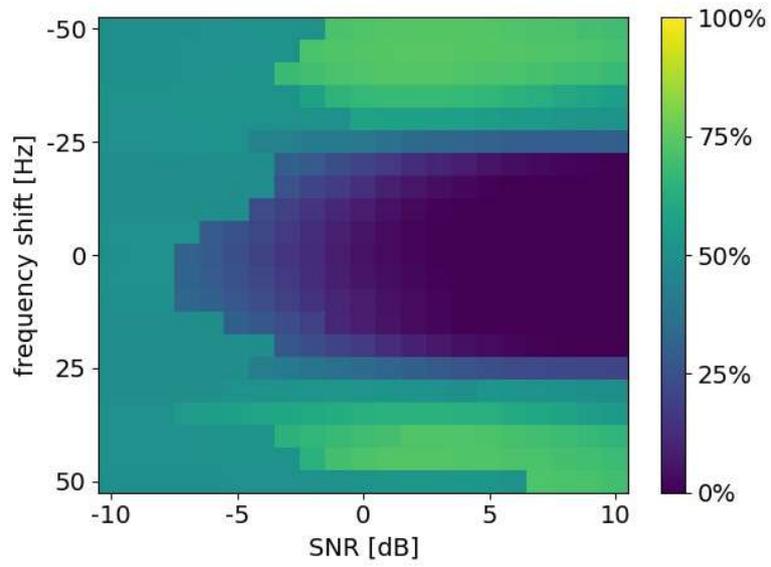


図 6.39 従来技術による SNR と周波数シフトの BER の関係

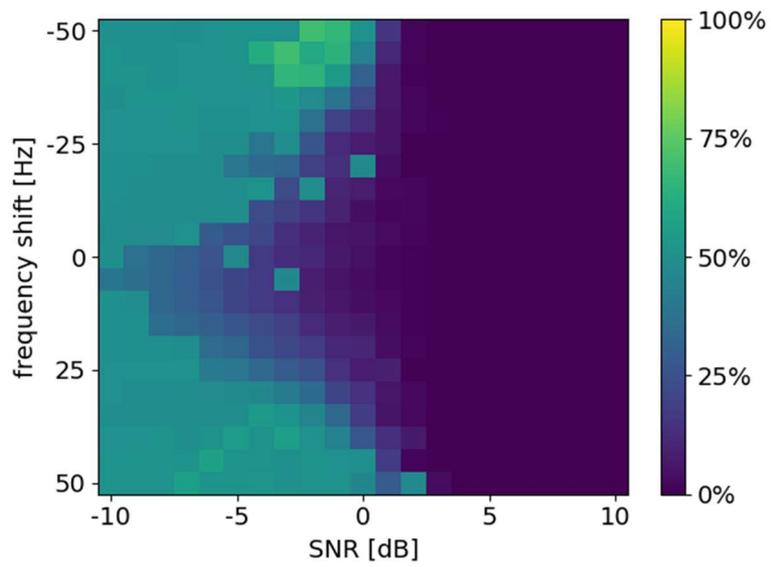


図 6.40 CNN による SNR と周波数シフトの BER の関係

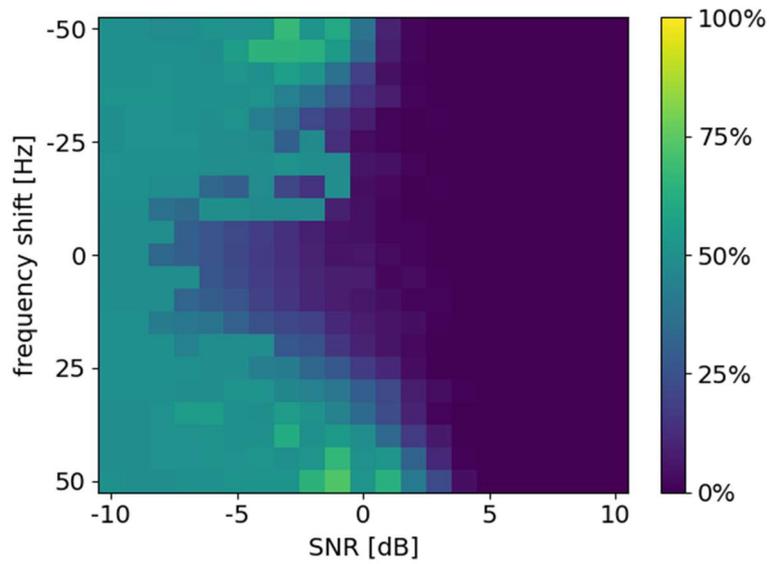


図 6.41 RNN による SNR と周波数シフトの BER の関係

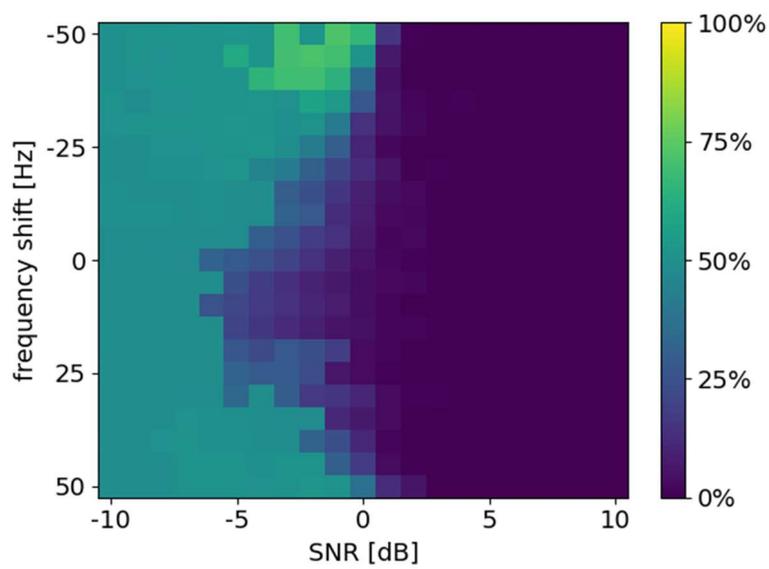


図 6.42 CRNN による SNR と周波数シフトの BER の関係

評価結果

図 6.37 の結果から従来および提案する超音波通信システムは、周波数シフトが発生しない場合は、SNR が 6dB 以上で BER は 0%を示している。また、今回提案する予測器は、雑音抑制や音源分離などの効果はないため、従来の超音波通信システムと精度は変わらない結果となった。一方、図 6.38 の結果から従来の超音波通信システムは、周波数シフトに対応ができないため同期処理の時点で誤りが生じ、BER は SNR を高くしても 50%に近い値となっている。これは、2 値の識別判定が全くできていないことを示している。一方、提案する超音波通信システムは、CNN・RNN・CRNN の 3 つの機械学習モデルを用いた予測器により、周波数シフトを含む音波信号からもバイナリ情報の取得に成功していることが分かる。特に、図 6.39 と図 6.40-6.42 の SNR と周波数シフトの BER の関係を比較すると、従来の超音波通信システムでは、25Hz~50Hz および-25Hz~-50Hz の範囲においてヒートマップの色はエラー率が高い黄色を示しているが、提案する超音波通信システムの各モデルによるヒートマップの色はエラー率が低い黒を示している。特に CRNN は高い周波数シフトが生じた場合でも 0dB 以上であれば、誤りが生じにくいことが明らかとなった。

考察

従来の超音波通信システムでは、周波数シフトの発生により SNR がよい環境下においても通信障害が発生してしまうことが重要な課題であったが、提案する超音波通信システムは、機械学習モデルによりスペクトログラムに含まれる周波数シフトの特性を学習することにより、周波数シフトの有無に関係なく同期処理および情報識別を高い精度で行うことが可能であることができたと考えられる。以上のことから機械学習モデルを用いた超音波通信システムの有効性が示されたと考えられる。

第7章

まとめと今後の展望

本研究では、非可聴域である 18kHz~20kHz の高周波帯域に情報を埋め込み通信を行う超音波通信システムにおいて、従来、情報識別・信号同期に誤り生じさせる原因であった環境雑音や反響音、残響音、ドップラー効果などの外乱に対して解析を行い、その問題に対して有効的な解決策を提案した。

従来の超音波通信システムでは、雑音環境下における識別精度は高い一方、周波数シフトの発生により SNR がよい環境下においても通信障害が発生してしまうことを確認した。そこで、本研究では、近年、環境音の解析や雑音抑制、音源分離などで発展している機械学習理論に注目し、音波信号に含まれる環境雑音や周波数シフトなどの外乱要素から情報識別・信号同期に必要な特徴量を抽出し機械学習に対応させることで、外乱環境下においても高い性能を示す超音波通信システムを設計することに至った。

機械学習モデルは通常、大規模なパラメータを持ち、学習や推論の処理に高い計算リソースを要求する。これは高い精度を達成する一方で、実際のシステムにモデルを組み込む場合、処理速度の観点からは課題が生じる。そこで、本研究では CNN, RNN, CRNN の3つの小規模な機械学習モデルを用い処理の違いによって外乱環境下における音波信号が持つ情報を効果的に取得するように学習を行った。実際の学習プロセスでは、白色雑音と周波数シフトにより劣化した音波信号を大量にシミュレーションし、時間変化と周波数成分の情報を2次元の画像として表現するスペクトログラムを用いることで、外乱下から音波信号の持つ情報を効果的に学習するよう設計した。

機械学習モデルを効果的に学習するには、大量のデータが必要であり、これには相応の計算リソースがかかる。そこで、学習プロセスの効率を向上させるために、まず初めに実際の音波信号よりもシンプルで計算効率の高い簡易音波信号モデルを設計し、CNN を用いた機械学習モデルの学習傾向を把握するに至った。そこで設計した簡易音波信号モデルでは、現実の複雑な音波信号を再現するよりも、基本的な特徴やパターンを抽出することに焦点を当てた。学習の結果は、学習曲線や未学習データを評価データとして用いて評価した。また、転移学習の適用により、シミュレーションにより生成したデータを学習した学習済みモデルに実際に録音した音波信号データを適用することで、モデルがシミュレーションデータから学習した知識を、現実の環境条件下において効果的に転移できるかを検証した。その結果、小規模な機械学習モデルであっても雑音環境下における精度向上が期待できることが分かった。また、シミュレーションにより作成した学習済みモデルを元に、ファインチューニングによる転移学習を行うことで、実環境が与える音波信号への歪や減衰の特徴を捉えることができ、識別精度の向上がみられた。

この簡易的な実験プロセスから機械学習モデルを設計する上での利点や課題点を考慮し、実際に超音波通信システムに用いる機械学習モデルを設計した。本研究で提案する超音波通信システムは、音波信号が BPF を経由する前に、その音波信号に含まれる周波数シフトを機械学習モデルで予測し、その予測結果を従来の BPF を用いた超音波通信システムに適応させることで、周波数シフトに対応する柔軟な超音波通信システムを提案するに至った。この提案手法の特色は、処理速度が遅い機械学習によるタスクを、同期処理および音波信号全体にかかる周波数シフトの予測に留め、出現回数を最小限とし、信号のバイナリ情報の予測には処理速度の速い BPF を用いた従来技術を用いることである。

提案する超音波通信システムでは、機械学習モデルによりスペクトログラムに含まれる音波信号のホッピングパターンを予測し、同期処理を行うホッピングパターン予測器と、周波数シフトの値を予測する周波数シフト予測器の、2つの予測器を用いるように設計した。

同期処理を行うホッピングパターン予測器は、機械学習モデルのパラメータ調整を行うために、グリッドサーチによる大規模な計算を行った。その結果、CNN は 81 モデル分、RNN は 54 モデル分、CRNN は 486 モデル分の学習を行った。また、周波数シフト予測器は、同期処理タスクの機械学習モデルにおいて最も高い精度のパラメータを用いて転移学習を行った。

従来技術を用いた超音波通信システムは、周波数シフトが発生しない場合は、SNR が 6dB 以上で BER は 0%を示したが、超音波通信システムに想定していないような周波数シフトが（本研究の場合は 25Hz 以上）発生すると 6dB 以上の SNR であっても BER が 20%以上上昇し、どれだけ SNR を高くしても周波数シフトによる BER は改善されないという、非常に重要な課題があったが、グリッドサーチにより調整された CRNN の機械学習モデルによる予測器は、SNR が 6dB 以上あるとき、周波数シフトが 0~100Hz まで発生しても誤ることなくホッピングパターンの検出が行われた。同様に CRNN の機械学習モデルによる周波数シフト予測器を用いると 6dB 以上の SNR であると、周波数シフトが発生した場合であっても BER が 0%を示した。

このことから従来の超音波通信システムでは、周波数シフトの発生により SNR がよい環境下においても通信障害が発生してしまうことが重要な課題であったが、提案する超音波通信システムは、機械学習モデルによりスペクトログラムに含まれる周波数シフトの特性を学習することができ、周波数シフトの有無に関係なく同期処理および情報識別を高い精度で行うことが可能であることが明らかとなった。

今後の展望として、提案した超音波通信システムがもたらす革新的な可能性に焦点を当てることが重要となる。第一に、この機械学習モデルに基づくシステムが実世界の複雑な環境下での実用性を確認する必要がある。異なるノイズや周波数変動が存在する状況においても、提案手法が高い信頼性で動作することが期待される。また、提案手法が通信の安定性向上にどの程度寄与できるかを定量的に評価し、他の通信手法と比較することも重要である。これにより、提案手法の優位性や限界がより具体的に理解され、実用段階での展開に向けた具体的な指針が得られると考えられる。

謝辞

本研究は、著者が北海道大学大学院情報科学科メディアネットワークコース修士課程時から開始し、北海道大学大学院情報科学科メディアネットワークコース博士後期課程も引き続き行ったものである。多大なるご指導を頂きました土橋宜典教授に心より感謝いたします。本論文をまとめるにあたり、副査をお引き受けいただいた北海道大学大学院情報科学研究所 メディアダイナミクス研究室 長谷山美紀教授，北海道大学大学院情報科学研究所 言語メディア学研究室 荒木健治教授，北海道大学大学院情報科学研究所メディア創生学研究室 坂本雄児教授に深く感謝いたします。本研究の学術的背景のご指導及び本研究の評価，方針に関して多大なるご助力をいただいた北海道大学大学院情報科学研究所情報メディア環境学研究室の青木直史助教に心より感謝いたします。

本研究の設計計画および実験方法について，貴重な御意見，御討論を頂きました(株)スマート・ソリューション・テクノロジー，池田 研一様，安田 寛様に心より深謝申し上げます。

博士後期課程在籍中の研究生活においてはアンビシャス博士人材フェロシップ制度(情報・AI)によるご支援をいただきましたものです。ここに深く感謝いたします。

著者の研究期間中，共同研究にご協力いただいた北海道大学大学院情報科学研究科メディアネットワークコース情報メディア環境学研究室の皆様心より御礼を申し上げます。

最後に，研究生活を支えてくださった家族に多大なる支援をいただきましたことに感謝の意を表し謝辞とさせていただきます。

参考文献

- [1] (株)スマート・ソリューション・テクノロジー, <https://www.sstinc.co.jp/>
- [2] “けいおんくん”,オムロンヘルスケア株式会社,
<https://www.healthcare.omron.co.jp/product/mc/mc-687.html>
- [3] “Another Track”,エヴィクサー株式会社, <https://www.evixar.com/technology>
- [4] G. Bang, M. Chung, and I. Ko, “Data communication method based on inaudible sound at near field”, ICACT ,pp. 639–642, 2016.
- [5] 西村 明, 坂本 真一, "音響データハイディングを用いるスピーカ再生音と同期した情報呈示システム," 電子情報通信学会論文誌, J93-A, no.2, pp.91-99, 2010.
- [6] 西村 明, " [招待講演] 音波通信技術とその評価," 信学技報, vol. 119, no. 163, EA2019-28, pp. 35-40, 2019.
- [7] 西村 明, " [招待講演] 空中音波通信技術とその応用",日本音響学会誌 77 卷 6 号 pp. 390–395, 2021.
- [8] 増野 淳, 藤野 洋輔, 工藤 理一, “6G 時代の多様な無線アクセスを支える先端無線技術の研究開発,” NTT 技術ジャーナル, 2022 年 5 月
- [9] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014, [online] Available:
<http://arxiv.org/abs/1409.1556>.
- [10] B. Shi, X. Bai and C. Yao, “An End-to-End Trainable Neural Network for Image based Sequence Recognition and Its Application to Scene Text Recognition,” IEEE Trans. Pattern Analysis and Machine Intel ligenge 2016
- [11] A. Jansson, et al., “SINGING VOICE SEPARATION WITH DEEP U NET CONVOLUTIONALNETWORKS,” Proceedings of the 18th ISMIR,2017
- [12] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, et al., "Deep Speech: Scaling up end-to-end speech recognition," arXiv:1412.5567, 2014.
- [13] W. Chan, N. Jaitly, Q. V. Le, et al., "Listen, Attend and Spell," arXiv:1508.01211, 2015.

- [14] D. Snyder, D. Garcia-Romero, G. Sell, et al., "Deep Speaker Embeddings for Diarization," arXiv:1705.02304, 2017.
- [15] X. Ma, X. Wang, C. Liu, et al., "Deep Residual Learning for Small-Footprint Text-Dependent Speaker Verification," arXiv:1705.02966, 2017.
- [16] A. van den Oord, S. Dieleman, H. Zen, et al., "WaveNet: A Generative Model for Raw Audio," arXiv:1609.03499, 2016.
- [17] Y. Wang, RJ Skerry-Ryan, D. Stanton, et al., "Tacotron: Towards End-to-End Speech Synthesis," arXiv:1703.10135, 2017.
- [18] 井本桂右, "音響イベントと音響シーンの分析," 音響学会誌, 74, 198–207 (2018).
- [19] 周藤 唯, 糸山 克寿, 西田 健次, 中臺 一博, "Mask U-Net を用いた環境音セグメンテーションの提案," 人工知能学会 AI チャレンジ研究会, 2018
- [20] Q. Jin, P. F. Schulam, S. Rawat, S. Burger, D. Ding and F. Metzger, "Event-based video retrieval using audio," Proc. Interspeech 2012, pp. 2085–2088 (2012).
- [21] Y. Koizumi, et al., "SNIPER: Few-shot Learning for Anomaly Detection to Minimize False-Negative Rate with Ensured True-Positive Rate," in Proc. of ICASSP, 2019.
- [22] E. Marchi, F. Vesperini, F. Eyben, S. Squartini and B. Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks," Proc. ICASSP 2015, pp. 1996–2000 (2015).
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need.," Advances in Neural Information Processing Systems, Vol. 30, pp. 6000–6010, 2017.
- [24] 小泉 悠馬, 安田 昌弘, 村田 伸, 齊藤 翔一郎, 原田 登, "SPIDERnet: ワンショット異常音検知のための深層ニューラルネットワーク," 日本音響学会 2020 年春季研究発表会, 2020.
- [25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," Proc.IEEE, pp.1–46, Nov., 1998
- [26] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations (ICLR), pp. 1-15
- [27] "SurfacePro7", Microsoft, <https://www.microsoft.com/ja-jp/d/surface-pro-7/92wrlrcvz4pr>
- [28] "Xperia ZX3", SONY, <https://www.sony.jp/xperia/xperia/softbank/xz3/>
- [29] "Brüel & Kjær Type 4231", Brüel & Kjær, <https://www.bksv.com/ja/transducers/acoustic/calibrators/sound-calibrator-4231>

- [30] “Brüel & Kjær Type 4190”, Brüel & Kjær,
<https://www.bksv.com/ja/transducers/acoustic/microphones/microphone-cartridges/4190>
- [31] 中山英樹, "深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習," 信学技報, vol. 115, no. 146, SP2015-45, pp. 55-59, (2015)
- [32] Machine Learning in Action, “Achieving 90% Accuracy in Object Recognition Task on CIFAR-10 Dataset with Keras : Convolutional Neural Networks,” 2018.
 (appliedmachinelearning.blog/2018/03/24/)
- [33] Zhen Zuo, Bing Shuai, Gang Wang, Xiao Liu, Xingxing Wang, Bing Wang, and Yushi Chen, “Convolutional recurrent neural networks: Learning spatial dependencies for image representation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 18–26.
- [34] K. Choi, G. Fazekas, M. Sandler, and K. Cho, “Convolutional recurrent neural networks for music classification,” arXiv preprint arXiv:1609.04243, 2016.
- [35] E. Cakır, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, “Convolutional recurrent neural networks for polyphonic sound event detection,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 6, pp. 1291–1303, June 2017.
- [36] K. Cho, B. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougres, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN encoder-decoder for statistical machine translation,” Proc. EMNLP 2014, pp. 1724–1734, Oct. 2014.
- [37] CHEN Yifu, 山田 武志, 牧野 昭二, “空間特徴と音響特徴を併用する音響イベント検出の検討”, 日本音響学会研究発表会講演論文集, ROMBUNNO.3-P-14, March, 2020.
- [38] 浅谷 尚希, 神谷 亨, 間普 真吾, 木戸 尚治, "2つの解像度スペクトログラムとTF-CRNN による呼吸音の分類", バイオメディカル・ファジィ・システム学会大会講演論文集, 2020, 33 卷, 33, p.64-67,
- [39] Q. Yang, Y. Zhang, W. Dai, and S.J. Pan, Transfer Learning, Cambridge University Press, 2020.
- [40] Jansson, A., Humphrey, E.J., Montecchio, N., Bittner, R., Kumar, A., Weyde, T.: Singing voice separation with deep U-Net convolutional networks. In: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), pp.323–332, 2017

- [41] H. Qian, M. Morise, M. Niitsuma, Y. Yamashita, “Development of a Communication System Using Sound Signals in Inaudible Frequency ands,” IPSJ Interaction 2015, pp.913–916, 2015. (in Japanese)

著者の研究業績

(A) 学術誌（学位論文関係）

[A-1]. Kosei Ozeki, Naofumi Aoki, Saki Anazawa, Yoshinori Dobashi, Kenichi Ikeda, and Hiroshi Yasuda, "Improving the Recognition Accuracy of a Sound Communication Systems Using a Neural Network," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2021, Volume E104.A, Issue 11, Pages 1577-1584, November 1, 2021.

(B) 国際学会（学位論文関係）

[B-1]. Kosei Ozeki, Naofumi Aoki, and Yoshinori Dobashi, "Development of an inaudible sound communications system based on a machine learning approach," IEEE The 2nd International Conference on Artificial Intelligence in Information and Communication 2020 (ICAIIIC2020), Fukuoka, Fukuoka, Japan, pp.79--83, February 19--21, 2020.

[B-2]. Kosei Ozeki, Naofumi Aoki, Saki Anazawa, Yoshinori Dobashi, Kenichi Ikeda, and Hiroshi Yasuda, "Designing a symbol classifier for inaudible sound communication systems using a neural network," International Symposium on Information Theory and its Applications (ISITA2020), Hawaii, USA, pp.569--572, October 24--27, 2020.

[B-3]. Kosei Ozeki, Naofumi Aoki, Saki Anazawa, Yoshinori Dobashi, Kenichi Ikeda, and Hiroshi Yasuda, "Evaluation on the Recognition Accuracy of an Inaudible Sound Communication System in Real Environments," 2020 International Conference on Emerging Technologies for Communications (ICETC 2020), P2-3, online, December 2--3, 2020.

[B-4]. Kosei OZEKI, Naofumi AOKI, Yoshinori DOBASHIA. "Acoustic event detection of ultrasonic signals using CNN", 2021 International Conference on Emerging Technologies for Communications (ICETC 2021), P2-7, online, December 2--3, 2021.

[B-5]. Kosei OZEKI, Naofumi AOKI, Yoshinori DOBASHIA. "Research on Machine Learning Model for Detecting Ultrasonic Signals", 2023 International Conference on Emerging Technologies for Communications (ICETC 2023), P3-24, Sapporo, Japan, November 29-- December 1, 2023.

(C)国内学会発表（その他）

- [C-1]. 尾関 剛成, 青木 直史, 土橋 宜典, "ニューラルネットワークによる非可聴域における音通信に関する一考察," 電子情報通信学会技術研究報告応用音響研究会資料, EA2019-1, 2019.7.
- [C-2]. 尾関 剛成, 青木 直史, 土橋 宜典, "機械学習による非可聴音通信システムの精度向上に関する一検討," 日本音響学会春季研究発表会, 埼玉大学 (さいたま), 2020.3. (スペシャルセッション)
- [C-3]. 尾関 剛成, 青木 直史, 土橋 宜典, "ニューラルネットワークによる音通信の識別率向上に関する一検討," 電気・情報関係学会北海道支部連合大会, 室蘭工業大学 (室蘭), 2019.11.
- [C-4]. 尾関 剛成, 青木 直史, 穴沢 早紀, 土橋 宜典, 池田 研一, 安田 寛, "機械学習による非可聴音通信システムにおける信号音識別に関する一検討," 電子情報通信学会技術研究報告応用音響研究会資料, EA2020-1, pp. 1-6, 2020.7.
- [C-5]. 尾関 剛成, 青木 直史, 土橋 宜典, "実環境における非可聴音通信システムの評価," 電気・情報関係学会北海道支部連合大会, オンライン開催, 2020.11.
- [C-6]. 尾関 剛成, 青木 直史, 穴沢 早紀, 土橋 宜典, 池田 研一, 安田 寛, "転移学習により設計された超音波信号識別器の評価," 日本音響学会春季研究発表会, オンライン開催, 2021.3.
- [C-7]. 尾関 剛成, 青木 直史, 土橋 宜典, 池田 研一, 安田 寛, "深層学習による超音波信号分類システムの開発," 信学技報, vol. 121, no. 66, SP2021-21, pp. 97-100, 2021.6.
- [C-8]. 尾関 剛成, 青木 直史, 土橋 宜典, "CNN を用いた超音波信号のイベント検出に関する検討," 電気・情報関係学会北海道支部連合大会, オンライン開催, 2021.11
- [C-9]. 尾関 剛成, 青木 直史, 土橋 宜典, 池田 研一, 安田 寛, "超音波信号の音響イベント検出システムの検討", 信学技報, vol. 122, no. 95, EA2022-22, pp. 47-50, 2022.7.
- [C-10]. 尾関 剛成, 青木 直史, 土橋 宜典, "音響イベント検出を用いた超音波通信システムの雑音下における性能評価", 日本音響学会, 北海道科学大学, 2022.9.
- [C-11]. 尾関 剛成, 青木 直史, 土橋 宜典, "超音波信号の検出を目的とした End-to-End モデルに関する検討", 電気・情報関係学会北海道支部連合大会, はこだて未来大学, 2023.10.

(D)受賞

- [D-1]. "2021 IEEE Sapporo Section Student Paper Contest, Best Presentation Award", IEEE Sapporo Section.2021