CF-018

お Approximate inverse model explanations を用いたパラ言語情報特定のための音響特徴抽出手法

Acoustic Sound Feature Extraction Methods for Paralinguistic Information Detection using Approximate Inverse Model Explanations

菅原 康史‡ 青木 慎太郎‡ 中城 裕之‡ 横山広樹‡ 中西 崇史‡
Yasufumi Sugawara Shintaro Aoki Hiroyuki Nakajo Hiroki Yokoyama Takafumi Nakanishi

1. はじめに

音声言語コミュニケーションでは，話し手と聞き手の間で様々な情報が伝達されている．その内，パラ言語情報は，話し手および聞き手において感情を想起させる重要な役割を担うと考えられる．パラ言語情報とは，話し手が聴き手への伝達を目的に意図的に表出する情報のうちイントネーション，リズム，声質などの韻律特徴によって伝達されることが多いために，文字に転写されることがないかまれである情報[1]のこととしている．

一般的に，パラ言語情報は，話し手が意図的に変化をつけることが可能であるとされている．その一方で，話し手が話し方につけた変化，例えば，プレゼンテーションの場面において日常会話の淡々とした話し方より抑揚を大きく付けリズム良く話すなどは，必ずしも話し手が意図した通りに伝わるとは限らない．加えて，パラ言語情報をもとにした客観的な評価は、聞き手に直接聞いてもらうことに依存している．また，得られる評価も，聞き手の主観的な側面が強いため，聞き手によって評価が異なることも稀ではない．

プレゼンテーション等のコミュニケーションにおいて，パラ言語情報を客観的に把握することが難しい．そのような場においてパラ言語情報を容易に抽出し，話し手にフィードバックすることができれば，より聞き手に伝わりやすいプレゼンテーション等のコミュニケーションを展開できると考える．

本稿では，Approximate Inverse Model Explanations (AIME) [2]を用いたパラ言語情報特定のための音響特徴抽出手法について示す．これまで，我々は AIME と呼ばれるブラックボックスモデル(AI，機械学習モデルなど)を対象として近似逆作用素を導出することにより，対局的，局所的特徴をそのモデルの挙動や入出力の説明として提示することが可能な手法である．本稿では AIME [2]を用いてパラ言語情報の手がかりとなる音響特徴抽出手法を示す．

本手法では，音声データからパラ言語情報を予測する機械学習モデル（ブラックボックスモデル）が存在すると仮定し，音声データから抽出した音響特徴の説明変数と，音声データの話し手が感情を込めている／感情を込めていない，の２クラスの目的変数を用いて，AIME による逆問題を解くことにより，パラ言語情報特定のための音響特徴を導出することが実現できると考えられる．

本研究により，音響特徴の中で，パラ言語情報の音響特徴を特定し，パラ言語情報の手がかりとなる音響特徴の可視化が実現すれば，話し手の抑揚やリズムなどのパラ言語情報に関連する音響特徴の変化がより明確に評価され，聞き手による客観的な印象のフィードバックを得ることが可能になると考えられる．これにより，話し手は自身のプレゼンテーション等の場面でのコミュニケーションスキルを，聞き手を介さずに客観的な評価や，改善するための具体的な指標を得ることができ，パラ言語情報の理解と応用の一助になると考えられる．

本稿の構成は以下の通りである．2 章では，XAI についてと音声特徴量抽出に関する関連研究について述べる.3 章では，本方式について述べる．4 章では，実験を行い，提案方式の有効性について検証する．5 章では，研究をまとめる．

2. 関連研究

本節では，本研究に関連する研究を紹介する．

AIME は，複雑な AI や機械学習モデルを「ブラックボックス」として扱い，その入出力データから近似的な逆演算子を生成して説明を導く，XAI（説明可能な AI）技術の一種である．AI や機械学習モデルの透明性，説明性，信頼性を高めるため，XAI の研究は多様な分野で進展しており，多くの論文が発表されている[2][3][4].

Speith[5]は XAI を大きく 2 つに分類している．1 つは，モデルの出力結果から説明を得る Post-hoc 手法であり[6]，もう 1 つはモデル自体が解釈可能である Ante-hoc 手法である[7]．Ante-hoc 手法には，線形回帰モデルや決定木，k-nearest neighbor モデル，ルールベースモデルなど，一般に説明可能とされる「グラスボックス」モデルが含まれる．これに対して，Post-hoc 手法は一般的にブラックボックスモデルに適用され，model-specific 手法と model-agnostic 手法に分かれる．Model-specific 手法は特定のブラックボックスモデルに限定されており，例えば Grad-CAM[8]は畳み込みニューラルネットワーク（CNN）に特化している．一方，model-agnostic 手法は LIME[3]，SHAP[4]，AIME[2]など，さまざまな AI モデルに適用できる汎用性がある．

model-agnostic 手法については 3 つの手法があると考える．1 つ目は，ブラックボックスの振る舞いを観察して説明を引き出す手法であり，partial dependence plots（PDP）[9][10]や individual conditional expectations（ICE）[11]がその例である．2 つ目は，順方向の手法を用いて重要な特徴を抽出する方法であり，LIME や SHAP がこれに該当する．3 つ目は，ブラックボックスの逆問題を解く手法であり，AIME がこのカテゴリに含まれる．

AIME の大きな特徴は，説明変数 X とその推定結果 Y があれば，ブラックボックスが存在しなくても近似逆作用素を導出できる点である．これにより，要因分析が可能となり，説明性を高めるための強力な手段となる．

津布久ら[12]は，ニュース記事から内閣支持率を予測するブラックボックスモデルを仮定し，AIME を用いて内閣

†武蔵野大学データサイエンス学部 Department of Data Science, Musashino University

支持率および不支持率の変動に寄与する単語を抽出する手法を提案した．本研究では，2018 年から 2023 年の「安倍首相」，「菅首相」，「岸田首相」に関するニュースデータを TF-IDF で解析し，内閣支持率のデータと統合して AIME を適用することで，内閣支持・不支持の変動に寄与する単語を特定した．この手法は内閣支持率の変動を引き起こす要素を定量的に明らかにするものであり．従来の XAI 手法よりもシンプルで解釈しやすい説明を提供する AIME の枠組みを活用することで，国民の反応をより正確に把握し，政策決定に活用できる可能性を示している．

Liu ら[13]は，パラ言語的特徴とスペクトラル特徴の抽出に基づいた，新たな音声感情分類システムを提案した．音声分類で一般的に用いられる，メル周波数ケプストラム係数(MFCC)をスペクトル特徴として抽出した．パラ言語特徴の抽出には，openSMILE[14]を使用した．これらの音声特徴を用いて，MLP および SVM モデルを訓練し，感情分類の性能を評価した．実験結果から，これらの機械学習モデルは、音声データにおける感情状態の分類において高い精度を示し，特にパラ言語特徴とスペクトル特徴の組み合わせが有効であり，複数の音響特徴と機械学習技術の統合的アプローチが有効であることを示した．

本研究では，音声コミュニケーションにおけるパラ言語情報の重要性とその客観的評価の困難さに着目し，これを解決するための手法を提案する．

パラ言語情報は，話し手の感情や意図を聞き手へと伝達する上で極めて重要であり，イントネーションやリズム，声質などの韻律特徴により伝達されるが，その評価は聞き手の主観的な要素が強く，より客観的評価が求められている．一方，AI 技術の進展に伴い， XAI 手法が注目されており，これをパラ言語情報の評価に応用することで，新たな知見を得ることが期待される．本研究は，XAI の一手法である AIME を用いて音響特徴を抽出し，パラ言語情報の特定および可視化を行うことを目指す．

AIME は，近似逆作用素を構築することで，モデルの出力結果に対する説明を提供するものであり，本研究ではこれを用いてパラ言語情報に関連する音響特徴を明確化する．本研究の位置付けとして，パラ言語情報の可視化と客観的評価を実現し，話し手の意図的な抑揚やリズムの変化がより正確に伝わり，評価される仕組みを提供する点が挙げられる．これにより，話者は自身の音声コミュニケーションスキルを客観的に評価することができ，パラ言語情報の理解と応用において重要な進展をもたらすと期待される．

3. AIME を用いたパラ言語情報特定のための音響特徴抽出手法

本節では，具体的な手法及び機能について示す．3.1 節では本方式における AIME の概要について述べる．3.2 節では，本方式の実現方法について述べる．また，本方式を構成する各機能の手法について述べる．3.3 節では AIME を用いた分析について述べる．

3.1 AIME 概要

AIME は，AI および機械学習モデルの挙動を解明するための説明可能な AI 技術の一つである．AIME の特徴は，ブラックボックスモデルを近似逆演算子を導出することによって説明する点にある．AIME の近似逆演算子を構築するためには，着目している AI，機械学習モデルの訓練データ

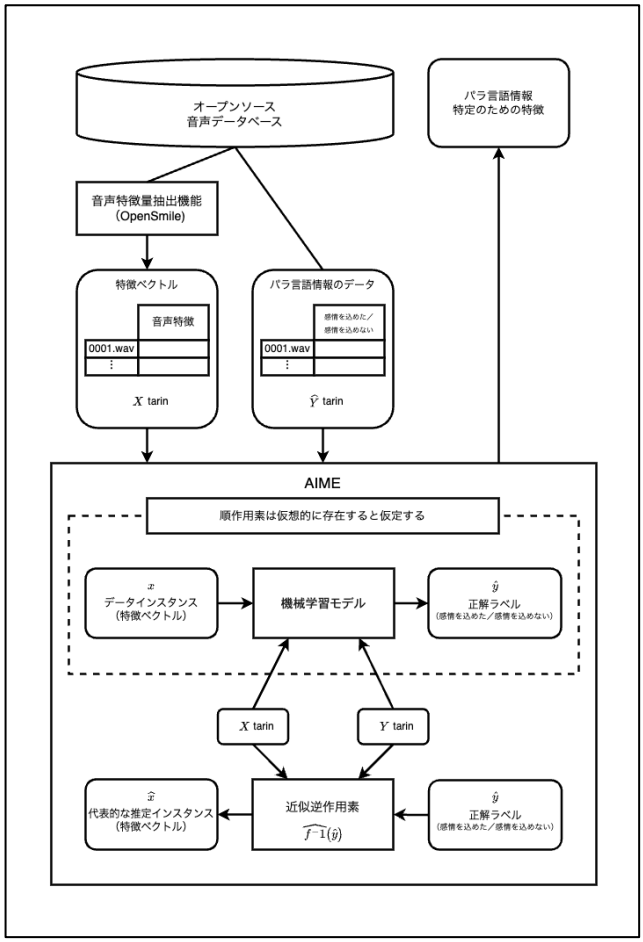

図 1　全体像

である X_train とその AI，機械学習モデルに X_train を入力して出力された Yˆを用いる．X_train は，音響特徴（本稿の場合 88 次元）×データ数で構成される行列である．Yˆは，2 クラス（感情を込めている/感情を込めていない)×データ数で構成される行列である．ここで，このような順作用素である AI，機械学習モデルが理想的な動作をすることを想定する．そのような場合，その理想的な動作をする順作用素は X_train を入力すると必ず Y_train をすることとなる．このような背景では，本方式では，文献[2]で提案された XAI のための AIME を拡張し，近似的な逆演算をするための AIME として，近似逆作用素を導出する際に X_train, Y_train を用いることとする．また，簡単化のため X_train を X, Y_train を Y と表現する．

このように，本方式のためにカスタマイズされた AIME は，ブラックボックスモデルの入出力データX, Yから，近似的な逆演算子を以下の式から導出される．

$$X = A^{\dagger}Y,$$
$$XY^T = A^{\dagger}YY^T,$$
$$XY^T(YY^T)^{-1} = A^{\dagger}(YY^T)(YY^T)^{-1},$$
$$A^{\dagger} = XY^{\dagger}(YY^T)^{-1} = XY^T$$

ここで上記の式より Y を Yˆに置き換える．また，Y^T は行列 Y の転置行列を示し，Y^†は Y のムーアペンローズの一般化逆行列を示す．これらの関係を図 2 に示す．

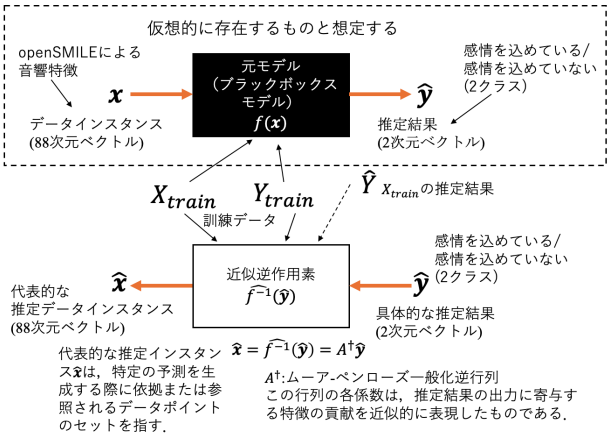

図２　本方式にカスタマイズされた AIME の概要

ここで， A^†は，音響特徴（本稿の場合 88 次元）×2 クラス（感情を込めている／感情を込めていない）で構成される行列となる．今回の場合，A^†は，各列は 2 クラス，つまり，感情を込めている／感情を込めていないを表し，それぞれに対して音響特徴がどのように貢献するかを示す対局的特徴重要度を示す．これにより，モデル全体の挙動を解釈することが可能となり，全体として，感情を込めた場合，感情を入れない場合のそれぞれの場合に・どの特徴量が予測に最も寄与しているかを明らかにする．

AIME は，対局的な特徴重要度を評価することにより，各特徴量が予測結果に与える影響を明確にし，モデルの全体的な挙動の把握が可能となる．これまでの対局的特徴重要度を導出手法と比較し，クラスごとの重要度を導出できる点，正/負の貢献をクリアに表現できる点が異なる．

3.2 システム全体像

本節では，AIME を用いたパラ言語情報特定のための音響特徴抽出手法の実現方法について示す．

本研究における提案手法の概要と機能群について述べ，本方式の全体図を図 1 に示す．

3.2.1 音声データベースについて

本節では，本研究で用いるオープンソースの音声データベースである No.7 音声データベース[15]について述べる．

No.7 音声データベースは， 研究用途・非商用での利用に限定して公開されている．話し手は，プロの日本人女性声優の小岩井ことりさんが担当しており，発話スタイルを分け，5 秒前後の収録した音声からデータベースが構成されている．また，音声データのファイル形式は WAV 形式であり，サンプリング周波数は 96 kHz，量子化ビット数は 24 bit となっている．

3.2.2 音声特徴抽出機能

本節では，音声ファイルから音声信号処理および特徴抽出のための標準的なツールである openSMILE を用いて音声特徴量を抽出する機能について述べる．

本機能では，No.7 音声データベースを用いて，音声ファイルごとに openSMILE を用いて音響特徴量を抽出する．また，openSMILE には複数の特徴量のセットが用意されており，本稿では eGeMAPSv01b という特徴量セットを利用し 1 つの音声ファイルに対して 88 個の特徴量を抽出する．

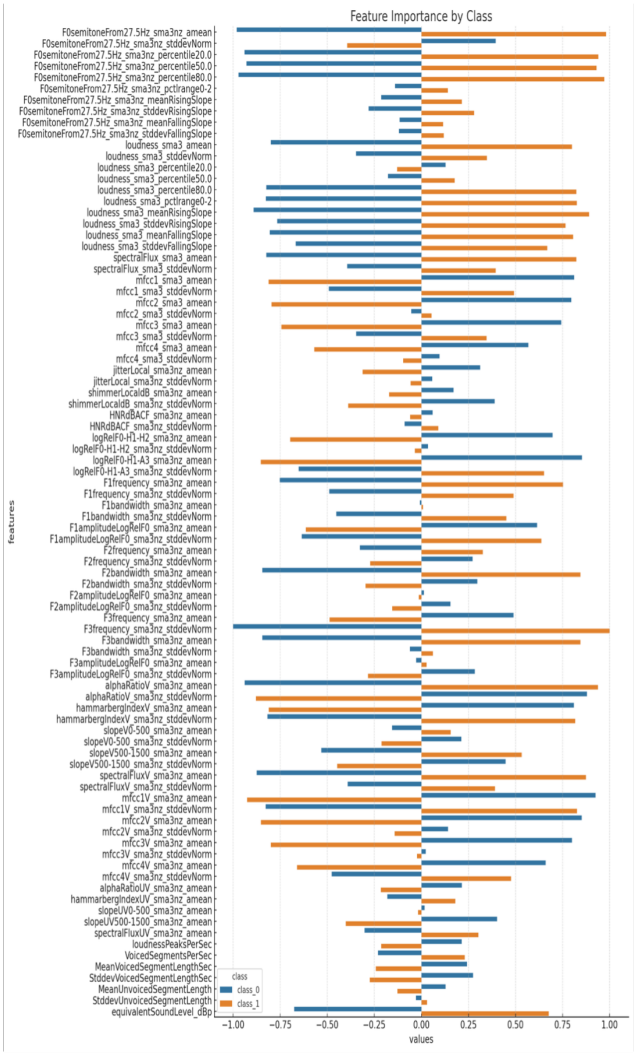

図３　実験１結果

3.3 AIME を用いた特徴抽出

近似逆行列$A^{\dagger}$を用いて，パラ言語特定に貢献する要因である特徴量を抽出する大局的特徴重要度を求めることが可能となる．本手法を用いて，パラ言語情報特定のための音響特徴抽出を行う．学習データとして研究者向け音声合成検証用 No.7 音声データベースから感情を込めている／感情を込めていないの 2 種類の音声データと，音声特徴抽出機能から抽出した 88 個の特徴量を用いて， AIME を用いて貢献している音声特徴量を特定する．

4 実験

本節では，AIME を用いたパラ言語情報特定のための音響特徴抽出手法の実験目的と概要について述べる．

4.1 節では，実験として AIME を用いて学習したモデルから評価に貢献する音声特徴量の対局的特徴重要度の結果について示す．4.2 節では，実験の出力結果をもとにした本方式の考察について示す．

4.1 実験 1 : AIME を用いた特徴抽出

本節では，実験 1 の結果を図 3 に示し、AIME を用いてパラ言語特定に貢献した音声特徴の大局的特徴重要度をも

とに貢献度が高い特徴量とそれらに関する説明について述べる．

結果から，F0semitone，loudness，spectralFlux，logRelF0-H1-A3 ， F2bandwidth ， F3frequency ， F3bandwidth ，alphaRatioV，hammarbergIndexV，mfcc1V，mfcc2V 等が高い貢献度として出力された．

以下は，出力された貢献度の高いそれぞれの特徴量に関する説明である．

- F0（Fundamental Frequency）は，ケプストラム法を用いて計算された基本周波数であり，音声信号の最も低い周波数成分の特徴量である．これは，話し手の声の高さという特徴を表している．
- Loudness は，聞き手の聴覚が感じる音の大きさを数値化したもので，音圧とは異なり，周波数が考慮された特徴量である．
- spectralFlux は，音声信号のパワースペクトルの変化量の尺度を表した特徴量であり，話し手の抑揚や声色の変化についての特徴を表しているとされている．
- logRelF0-H1-A3 は第 3 フォルマントの最高倍音である A3 に対する F0 の最低倍音である H1 の比の値である．話し手の緊張度や息漏れなどを表す数値として扱われることのある特徴量である．また，フォルマントとは，話し手の発音などで，特定の周波数で増幅される周波数領域のことであり第 3 フォルマントは，複数ある周波数領域の低い順から，3 番目の周波数領域を指す．
- F2frequency は，第 2 フォルマントと呼ばれ，声道の長さによって異なるが，約 1500Hz の周波数帯域を指す． F2frequency は，複数ある周波数領域の低い順から，２番目に位置する周波数領域であり，母音の音色や発話明瞭度，母音の識別などに用いられる特徴量である．また，openSMILE を用いて抽出した特徴量には，frequency と bandwidth の 2 つがある．これは，frequency は特定の周波数を，bandwidth は特定の周波帯を表している．
- F3frequency は，第 3 フォルマントと呼ばれ，声道の長さによって異なるが，約 2500Hz の周波数帯域を指し，F2frequency と同様に，母音の音色や発話明瞭度，母音の識別などに用いられる特徴量である．
- Alpha Ratio は，声帯振動における低周波数域帯（50～1000Hz）と高周波数域帯（1～5kHz）の合計エネルギー比である．音質や声の明瞭さを評価するために用いられる特徴量である．
- Hammarberg Index は，声帯振動における低周波数帯（(0～2kHz)と高周波数域帯（2～5kHz）の最大エネルギー比である．こちらも Alpha Ratio と同じく，音質や声の明瞭さを評価するために用いられる特徴量である．
- MFCC は，メル周波数ケプストラム係数とも呼ばれ，音声信号を短時間フレームに分割し、各フレームにフーリエ変換、メルフィルタバンク、対数変換、離散コサイン変換を順に適用して算出されたものである．聞き手の聴覚特性を考慮している特徴量である．

また，openSMILE を用いて抽出した各特徴量にはそれぞれ amean（平均）と stddevNorm（正規化された標準偏差）の 2 つ統計量が存在する．ameans は全体的な傾向を捉えるのに対し stddevNorm はデータのばらつきを示している．音声特徴の平均値は全体的なトレンドや代表的な値を示し，特定の区間における感情や意図の変化を直接的に捉えることができる．それに対して，正規化標準偏差は特定の区間における値の変化を表しており，全体的なトレンドを表す指標とは言えない．そのため，平均値を用いた特徴量が高い重要度を示す一方で，正規化標準偏差は相対的に重要度が低くなる傾向があると考えられる．具体例をひとつ挙げると， mfcc1_sma3_amean，mfcc1V_sma3nz_stddevNorm はそれぞれ MFCC の平均値，正規化標準偏差を示しているが，正規化標準偏差は平均値に比べて重要度が低くなっている．F0 や Loudness など他の特徴量でも同様に平均値は重要度が高くなる傾向が見られる一方で，標準偏差は瞬間的な変化を捉えることはできるものの，全体的なトレンドを示すには不安定であるため相対的に重要度は低くなる．これらの結果から，AIME を用いた音響特徴抽出においては，平均値を用いた特徴量がパラ言語情報の特定において有効であることが示唆される．

4.2 実験考察

本節では，実験から得られた特徴量をもとに，提案手法の有効性を考察する．

まず， F0（Fundamental Frequency），loudness，MFCC の 3 つの特徴量について考察する．この３つの特徴量は，音声による感情認識モデルや音声認識モデルにも用いられることのある特徴量である．それぞれ，話し手の声の高さ，声の大きさ，聞き手の聴覚特性を考慮した話し手の特徴を表している．これらに関しては，総じて話し手の感情の強弱や意図の強調を表現する上で不可欠な要素であり，聞き手にとっても話し手の感情や意図の理解に必要な要素であることが示唆される．そのため，この 3 つの特徴量はパラ言語情報特定のための音響特徴として有用であると考えられる．

次に，spectralFlux について考察する．この特徴量は，音声信号のスペクトルの変化量であり，話し手がつけた抑揚や声色の変化を捉えることが可能である．話し手の抑揚や声色の変化は，感情の表現や意思伝達の強調において重要な役割を果たすと考えられる．そのため，spectralFlux もパラ言語情報特定のための音響特徴として有用であると考えられる．

次に，F2frequency と F3frequency の 2 つの特徴量について考察する．この 2 つの特徴量は，母音に大きく関係している．一般的に，話し手の母音の発音が曖昧であると，単語が聞き取りづらくなり明瞭さが低くなる．また，話し手が母音をはっきりと発音することによって，聞き手への感情やメッセージのニュアンスの伝達がより確実なものになる．つまり，話し手の母音の強調は，話し手自身の感情の表現力が増し，聞き手の感情を想起させるという点において，パラ言語情報であると捉えることができる．そのため，F2frequency と F3frequency もパラ言語情報特定のための音響特徴として有用であると考えられる．

次に，alpha Ratio と hammarberg Index の 2 つの特徴量について考察する．この 2 つの特徴量は，特定の周波数帯の

エネルギー比を算出したもので，話し手の音質や声の明瞭さを表すとされている．この 2 つの特徴量は，F2frequency と F3frequency と関連しており，話し手がつけた母音や子音の発音の強弱が、声帯振動の強弱となるため、その強弱がエネルギー比としてあらわれる． そのため，alpha Ratio と hammarberg Index もパラ言語情報特定のための音響特徴として有用であると考えられる．

次に，logRelF0-H1-A3 について考察する．この特徴量は，話し手の緊張度や息漏れを表すものであり，聞き手に対して直接的に感情を想起させるような要素でないと考えることができる．しかし，話し手が適度な緊張感のもと軽快なリズム感で息継ぎをするという話し方は，聞き手に対して良い側面の影響があると考えられる．そのため，logRelF0-H1-A3 はパラ言語情報特定のための音響特徴として有用であるとは言い切れないものの，全く有用ではないとも言えないと考えられる．

一方で，相関が見られなかった特徴量も存在する．例えば，jitterLocal（基本周波数のゆらぎ），shimmerLocal（振幅のゆらぎ），slopeV500-1500（500Hz から 1500Hz の間のスペクトル傾斜），meanUnvoicedSegmentLength（無声音の平均長さ）などは，相関が低かった．これらの特徴量がパラ言語情報の特定において貢献しなかった理由を考察する．まず，jitterLocal や shimmerLocal は，声の安定性や微小な振幅変動を示すが，これらはパラ言語情報を伝える上での主要な指標とは言えない．これらの特徴量は，主に音声の質や声帯の健康状態を評価するためのものであり，感情や意図の伝達には直接的に関与しないと考えられる．

また，slopeV500-1500 は特定の周波数帯域におけるスペクトルの傾斜を示すが，この帯域の情報が感情や意図の伝達においては，特に 500Hz から 1500Hz の範囲は，人間の聴覚が感情の微妙な変化を捉えるのにはあまり寄与しないと考えられるため，重要ではない可能性がある．さらに，meanUnvoicedSegmentLength は無声音の平均長さを示すが，無声音部分は感情や意図の伝達において重要な役割を果たさないため，この特徴量が低い相関を示したのは自然である．

これらの結果から，パラ言語情報の特定には，単純な物理的特性や声の安定性だけではなく，声の高さや強弱，スペクトルの動的な変化といった複雑な音響特徴が重要であることが明らかになった．この知見は，音声コミュニケーションの質を向上させるための新たなアプローチを提供するものである．また，今回の研究結果は，感情や意図を伝えるためには，より動的で感情に富んだ音響特徴量が必要であることを示唆している．これにより，話し手の感情状態や意図を正確に把握するためには，単純な平均値や分散だけでなく，音声信号の変動や動きに注目する必要があることが示された．

5 終わりに

本研究では，AIME を用いたパラ言語情報特定のための音響特徴抽出手法を提案し，その有効性を検証した．パラ言語情報は，話し手の感情や意図を伝達する上で重要な役割を果たすが，従来の手法ではその客観的評価が困難であった．本研究は，機械学習モデルの出力を説明する AIME を活用し，音響特徴を抽出することで，パラ言語情報の特定および可視化を実現した．

実験結果から，F0，loudness，spectralFlux，F2frequency，F3frequency，alpha Ratio，hammarberg Index といった音響特徴量がパラ言語情報の特定に有効であることが示された．これらの特徴量は，話し手の感情や意図を明確に伝えるために重要であり，特にプレゼンテーションなどの場面において抑揚やリズムを評価するための有力な指標となる．

さらに，本研究の重要な発見として，話し手の発言が意識的に行われたものか無意識的に行われたものかを区別する必要性が明らかになった．この区別が重要である理由は，意識的な発言と無意識的な発言では，聞き手に与える影響や意図の伝達に大きな違いがあるためである．意識的な発言は，話し手が意図的に強調したいポイントや感情を伝えるための手段として用いられることが多い．一方で，無意識的な発言は，話し手の自然な感情や思考の表出を反映しており，これもまた重要な情報である．このため，両者を区別することで，より精緻なコミュニケーション分析が可能となり，音声コミュニケーションの質を高めることができる．

今後の展望として，上記課題を解決することで，話し手の意図をより精緻に捉えることが可能となり，音声コミュニケーションの質を飛躍的に向上させることが期待される．特に，教育やビジネスの分野において，パラ言語情報の客観的評価は重要な意味を持つ．本手法の実用化により，プレゼンテーションスキルの向上やコミュニケーションの円滑化が図られることが期待される．

本研究は，パラ言語情報の評価における新たな視点を提供し，音声コミュニケーションの理解と改善に貢献するものである．今後もこの分野での研究を進め，発話者の感情の有無だけでなく，その発言が意識的に行われたのか無意識的に行われたのかを明らかにするための新たな手法の開発とその応用を追求する．これにより，話し手と聞き手の間でより正確で深いコミュニケーションが可能となることを目指す．

参考文献

[1] 鈴木 美穂，"音声コミュニケーションにおける感情表現とパラ言語情報の役割"，日本音響学会誌，Vol.66，pp.139-147 (2010).

[2] Nakanishi Takafumi, "Approximate Inverse Model Explanations (AIME): Unveiling Local and Global Insights in Machine Learning Models", in IEEE Access，Vol.11，pp.101020-101044 (2023).

[3] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin，"Why should I trust you? Explaining the predictions of any classifier" , Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016).

[4] Scott M. Lundberg, and Su-In Lee, "A unified approach to interpreting model predictions" Advances in Neural Information Processing Systems，Vol.30 (2017).

[5] Timo Speith, "A review of taxonomies of explainable artificial intelligence (XAI) methods" in: Association for Computing Machinery, New York, NY, USA， pp.2239-2250 (2022).

[6] Amina Adadi, Mohammed Berrada, "Peeking inside the black-box:A survey on explainable artificial intelligence (XAI)", IEEE access, Vol.6, pp.52138-52160 (2018).

[7] Mengnan Du, Ninghao Liu, Xia Hu, "Techniques for interpretable machine learning", Communications of the ACM，63(1)，pp.68-77 (2019).

[8] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization." , Proceedings of the IEEE International Conference on Computer Vision, pp.618-626(2017).

[9] Jerome H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Statist., Vol. 29，No. 5，pp. 1189-1232 (2001).
[10] Qingyuan Zhao, Trevor Hastie, "Causal interpretations of black-box models," J. Bus. Econ. Stat., Vol. 39，No. 1， pp. 272-281 (2017).
[11] Alex Goldstein, Adam Kapelner, Justin Bleich, Emil Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," J. Computat. Graph. Stat., Vol.24，No.1，pp.44-65 (2015).
[12] 津布久尚貴，峰松彩子，岡田龍太郎，中西崇文．: "Approximate Inverse Model Explanations (AIME)を用いた内閣支持率変化要因抽出手法"，DEIM (2023).
[13] Tong Liu, Xiaochen Yuan, "Paralinguistic and spectral feature extraction. for speech emotion classification using machine learning techniques." EURASIP Journal on Audio，Speech and Music Processing (2023).
[14] Florian Eyben, Martin Wöllmer, Björn Wolfgang Schuller, "openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor," in Proc. of the 18th ACM International Conference on Multimedia，pp. 1459-1462 (2010).
[15] 森勢 将雅，"声を含むデータベースの「使いやすさ」に関する一考察～No.7 音声・歌唱データベース構築を実例として～"，情報処理学会音声言語情報処理研究会，Vol. 2022-SLP-144，No. 20， pp. 1-6(2022).